

**ENCYCLOPEDIA OF
HUMAN INTELLIGENCE**

EDITORIAL BOARD

Editor in Chief

Robert J. Sternberg
Yale University

Editors

Stephen J. Ceci
Cornell University

John Horn
University of Southern California

Earl Hunt
University of Washington

Joseph D. Matarazzo
Oregon Health Sciences University

Sandra Scarr
University of Virginia

421
.E
177

ENCYCLOPEDIA OF HUMAN INTELLIGENCE

Robert J. Sternberg

Editor in Chief

MACMILLAN PUBLISHING COMPANY
NEW YORK

Maxwell Macmillan Canada
TORONTO

Maxwell Macmillan International
NEW YORK OXFORD SINGAPORE SYDNEY

DAYTON MEMORIAL LIBRARY
REGIS UNIVERSITY

REGIS UNIVERSITY
LIBRARY

MAR 15 1995

RECEIVED

Copyright © 1994 by Macmillan Publishing Company
A Division of Macmillan, Inc.

All rights reserved. No part of this book may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without permission in writing from the Publisher.

Macmillan Publishing Company
866 Third Avenue, New York, NY 10022

Maxwell Macmillan Canada, Inc.
1200 Eglinton Avenue East, Suite 200, Don Mills, Ontario M3C 3N1

Macmillan, Inc., is part of the Maxwell Communication Group of Companies.

PRINTED IN THE UNITED STATES OF AMERICA

printing number

1 2 3 4 5 6 7 8 9 10

Library of Congress Cataloging-in-Publication Data

Encyclopedia of human intelligence / Robert J. Sternberg, editor in chief.

p. cm.

Includes bibliographical references and index.

ISBN 0-02-897407-7 (set : alk. paper)

1. Intellect—Encyclopedias. 2. Intelligence levels—Encyclopedias. I. Sternberg, Robert J.

BF431.E59 1994

153.9'03—dc20

93-46975

CIP

The paper used in this publication meets the minimum requirements of American National Standard for Information Sciences—Permanence of Paper for Printed Library Materials. ANSI Z39.48-1984.

CONTENTS

Preface	ix
List of Articles	xi
List of Contributors	xix
ENCYCLOPEDIA OF HUMAN INTELLIGENCE	1
Index	1177

Editorial and Production Staff

Philip Friedman

Publisher

Elly Dickason

Editor in Chief, Macmillan Reference

David Eckroth

Executive Editor

Ann Kathleen Bradley

Project Editor

Sabrina Bowers

Editorial Assistant

Leland S. Lowther Donald Spanel

Copy-editors

Robert Tennenhouse

Proofreader

Terri Dieli

Production Manager

PREFACE

What is intelligence? What is the relation of IQ to intelligence? Why are some people more intelligent than others? To what extent is intelligence inherited and to what extent is it a product of the environment? Can people increase their intelligence? These are just a handful of the many questions addressed in the *Encyclopedia of Human Intelligence*, the most comprehensive and definitive compendium of information about human intelligence ever published.

This two-volume work contains over 250 articles of varying lengths dealing with all aspects of human intelligence. The articles have been written and edited by the foremost scholars in the world in the field of human intelligence. These volumes will be informative to *anyone* with an interest in intelligence—students, parents, teachers, and other professionals, and even experienced psychologists. The volumes truly contain something for everyone.

The titles of the articles convey the range of topics included in the work—topics such as Ability Grouping, Aging and Intelligence, Artificial Intelligence, Birth Order, Creativity, Criminality, and Culture, to name just a few topics from the As, Bs, and Cs. An international team of contributors makes this book of interest not only in the United States but in countries around the world.

Human intelligence is one of the most important fields of study, in psychology as in any other discipline. Arguably, it is our most important “natural” resource. What the future holds for us is limited only by our drive and our intelligence. By browsing through these volumes, everyone can learn what we know about intelligence and its role in science, education, and society.

Because this book is an encyclopedia, it can be read in various ways. One way, obviously, is to turn to those articles that are of immediate interest. The alphabetical order of the entries renders each topic easy to find, and cross-referencing makes it a simple matter to learn more about related topics. Words set in small capital letters refer the reader to an article by that title. A second way to read this work is from start to finish. Read in this way, the volumes will provide a thorough course on the topic of human intelligence. A third way to read is by themes. For example, there are multiple articles related to testing, education, job performance, and the like. However the volumes are read, they will satisfy the reader with even the most voracious appetite for information on human intelligence.

I am grateful to my associate editors and to the authors of the articles for their hard work in producing each article. I am also grateful to the editors at Macmillan Publishing

Company who have helped give birth to these two volumes—Elly Dickason, David Eckroth, and Ann Bradley. Finally, I thank you, the reader, for your interest in this intriguing field and for learning about what it is that makes the study of human intelligence a field of unending fascination.

ROBERT J. STERNBERG

LIST OF ARTICLES

A

- Abecedarian Project
Craig T. Ramey
- Abilities and Aptitudes
Richard E. Snow
- Ability Grouping
Karen B. Rogers
- Abstraction
Patrick C. Kyllonen
- Achievement Testing
Michael Zieky
- Adaptive Behavior
Henry Leland
- Adaptive Behavior, Assessment of
Mary diSibio
- Adulthood, Achievement in
Richard Wagner
Rufus L. Carter
- African Americans
Melvin N. Wilson
- Age and the Content of Intelligence Tests
Bert Hayslip, Jr.
- Aging and Intelligence
Ulman Lindenberger
Paul B. Baltes
- Agnosia
Carol J. Schramke
- Alcohol and Alcohol Abuse
Sara Jo Nixon
Oscar A. Parsons
- American Association on Mental Retardation
Pamela Woodley-Zanthos
Alfred A. Baumeister
- American College Test
James Maxey
- Anastasi, Anne
Florence Denmark
- Animal Intelligence: Historical Perspectives and Contemporary Approaches
Louis M. Herman
Adam A. Pack
- Animal Intelligence: Primate
Duane M. Rumbaugh
David A. Washburn
- Anxiety
Charles D. Spielberger
Sumner J. Sydeman
- Aphasia
George W. Hynd
Scott S. Rubin
- Aptitude Tests
Robert J. Gregory
- Aptitude-Treatment Interaction
Richard E. Snow
- Armed Services Vocational Aptitude Battery
Gerald Larson
- Army Alpha and Beta Tests of Intelligence
Frederick L. McGuire
- Artificial Intelligence
Roger Schank
- Artistic Ability: Children's Drawings
Claire Golumb

Asian Americans

Lisa A. Suzuki

Terry B. Gutkin

Athletic Ability

Daniel M. Landers

Attention

Steven Yantis

Attention Deficit Hyperactivity Disorder

Mark F. Ward

Auditory Abilities

Lazar Stankov

Autism

Fred R. Volkmar

Donald J. Cohen

B

Bayley, Nancy

Lewis P. Lipsitt

Behaviorism

Robert Epstein

Benton, Arthur L.

Kerry deS. Hamsher

Bias in Testing

Cecil R. Reynolds

Bilingualism

Kenji Hakuta

Binet, Alfred

Joseph D. Matarazzo

Bioecological Theory of Intelligence

Stephen J. Ceci

Biological Measures of Intelligence

Joseph D. Matarazzo

Birth Defects

Joan C. Martin

Birth Order, Spacing, and Family Size

Joseph Lee Rodgers

David Rowe

Bond Sampling Theory of Human Abilities

William C. Tirre

Brain

Arthur P. Shimamura

Joan W. Shimamura

Brain, Pathologies of the

Arthur Benton

Brigham, Carl C.

William H. Angoff

Broca, Paul

Manfred Meier

David Kent Mercer

Burt, Cyril L.

Sandra Scarr

C

Carroll, John B.

Lyle V. Jones

Case Reports

Alan S. Kaufman

Nadeen L. Kaufman

Cattell, James McKeen

Robert L. Solso

Cattell, R. B.

Paul Kline

Child and Adolescent Intelligence

Charles J. Brainerd

Chromosomal Abnormalities

Joseph H. Hersh

Allan Bloom

Margaret J. Barch

Classification of Intelligence

Robert J. Gregory

Complexity, Cognitive

Peter Suedfeld

Cognitive Styles

Nathan Kogan

Cohort Effects

Christopher Hertzog

College Board

Dean K. Whitla

Compensatory Mechanisms

Roger A. Dixon

Lars Bäckman

Competence Versus Performance

Susan A. Gelman

Connectionism

Mitchell Rabinowitz

Contextualist Theories of Intelligence

Daniel P. Keating

Creativity

Mihaly Csikszentmihalyi

Crime and Delinquency

Lisabeth F. DiLalla

Lorrie A. Nielson

Criminality

James R. P. Ogloff

David G. Schaefer

Cronbach, L. J.

Richard E. Snow

Cross-Cultural Variations in Intelligence

John W. Berry

Culture-Fair and Culture-Free Tests

Sandra Scarr

Culture and Intelligence

John U. Ogbu

D

Decision Making and Judgment

Harriet Shaklee

Development, Cognitive

Robert S. Siegler

Deviation Quotient

Joseph D. Matarazzo

Dialectical Thinking

Gisela Labouvie-Vief

Differential Ability Scales

Mark H. Daniel

Doll, Edgar A.

Richard T. von Mayrhauser

Down Syndrome

Curtis K. Deutsch

Draw-a-Figure Test

Patti L. Harrison

Hannelore Schock

Drugs and Intelligence

Gary L. Wenk

Dynamic Assessment of Mental Abilities

Jerry S. Carlson

Dyscalculia

Wiley Mittenberg

Dyslexia

Ingrid N. Leckliter

E

Eastern Views of Intelligence

J. P. Das

EEG Evoked Potentials

Hans J. Eysenck

Error of Measurement

Edward J. Slawski

Ethnicity, Race, and Measured Intelligence

Lisa A. Suzuki

Damian A. Vraniak

Evolution of Human Intelligence

Lisabeth DiLalla

Carol Patrick

Experimental Design

David Rindskopf

Eysenck, Hans J.

Arthur R. Jensen

F

Facet Theory

Ingwer Borg

Factor Analysis

John J. McArdle

Factor Analysis of the Wechsler and Stanford-Binet Scales

Alan S. Kaufman

Family Environments

David C. Rowe

Joseph L. Rodgers

Fetal Alcohol Syndrome

Heather Carmichael Olson

Fluid and Crystallized Intelligence, Theory of

John L. Horn

Fluid and Crystallized Theory of Intelligence, Measures of

Richard W. Woodcock

G

Galton, Francis

Arthur R. Jensen

Gender Differences in Intellectual Abilities

Diane F. Halpern

General Aptitude Test Battery

John E. Hunter

General Intelligence
Jan-Eric Gustafsson

Genetics, Behavior
John C. Loehlin

Genius
Michael J. A. Howe

Gesell, Arnold
Louise Bates Ames

Giftedness
Frances Degen Horowitz

Gifted and Talented Students, Evaluation of
Educational Programs for
David M. Fetterman

Glaser, Robert
Lauren Resnick

Goddard, Henry H.
Henry Leland

Goldstein, Kurt
Arthur Benton

Group Tests
John Fremer

Guilford, J. P.
Philip R. Merrifield

Guttman, Louis E.
Ruth Guttman

H

Halstead, Ward Campbell
Ralph M. Reitan

Health and Intelligence
Howard S. Friedman

Hebb, Donald O.
Colin M. MacLeod

Hebb's Theory of Intelligence
Douglas A. Bors

Heritability
Peter H. Schönemann

Hierarchical Theories of Intelligence
Jan-Eric Gustafsson

Hispanics
Lisa A. Suzuki
Terry B. Gutkin

Hollingworth, Leta S.
Julian C. Stanley

Humphreys, Lloyd G.
Phillip L. Ackerman

Hunt, Joseph McVicker
Ina Č. Užgiris

Hyperactivity
Mark F. Ward

I

Illiteracy
Jeanne S. Chall
Mary E. Curtis

Imagery
Ronald A. Finke
Jennifer J. Freyd

Individual Tests
Jerome M. Sattler
Naomi G. Singer

Infancy
Marc H. Bornstein

Infant Tests as Measures of Early Competence
Henry N. Ricciuti

Information Processing
Patrick C. Kyllonen

Insight
Janet E. Davidson

Intelligent Quotient
Joseph D. Matarazzo

Interactionist Views on Intelligence
Jaan Valsiner

Interference
Frank N. Dempster

Intervention, Infant and Preschool
Alice Sterling Honig

Interventions, Later
Raymond S. Nickerson

Intuition
Kenneth S. Bowers

IQ Gains over Time
James R. Flynn

J

Japanese
Lisa A. Suzuki
Terry B. Gutkin

Jensen, Arthur R.

Cecil R. Reynolds

Job Performance

Malcolm Ree

K

Kaufman Assessment Battery for Children

Cecil R. Reynolds

Knowledge

Kenneth J. Gilhooly

L

Language and Intelligence

Julie A. Eisele

Dorothy M. Aram

Latent Trait Theory

Susan Embretson

Paul DeBoeck

Learning Disability

Antonia A. Forster

Learning and Intelligence

Earl C. Butterfield

Learning Potential Assessment Device

Feuerstein, Reuven

Learning, Skill, and Transfer

Phillip L. Ackerman

Legal Issues in Intelligence

Paul J. Hofer

Localization of Brain Function

Robert A. Bornstein

Anthony M. Podraza

Luria, A. R.

J. P. Das

M

Mainstreaming

Gary Bunch

Mathematical Ability

Richard E. Mayer

McCarthy Scales of Children's Abilities

James S. Gyurke

Measurement and Prediction of Intelligence

Lloyd G. Humphreys

Mechanical Ability

David L. Alderton

Memory

Randall W. Engle

Memory, Exceptional

K. Anders Ericsson

Mental Age

Joseph D. Matarazzo

Mental Measurements Yearbook

Joseph D. Matarazzo

Mental Retardation, Cultural-Familial

Robert M. Hodapp

Mental Retardation, Organic

John M. Belmont

Mercer, Jane R.

Cecil R. Reynolds

Beth Murdock

Metacognition

Earl C. Butterfield

Motivation

Carrol E. Izard

Brian P. Ackerman

Motor Ability

Lewis P. Lipsitt

Multiple Intelligences Theory

Howard Gardner

Musical Ability

Diana Deutsch

Musical Intelligence

Lyle Davidson

Bruce Torff

N

Native Americans

Damian A. Vraniak

Nature, Nurture, and Development

Robert Plomin

Neuropsychology, Clinical

Ralph M. Reitan

Deborah Wolfson

Norms

Richard W. Woodcock

Nutrition

Henry N. Ricciuti

O

Occupations

*Rene Dawis***P**

Parenting and Intelligence

Wendy M. Williams

Perception

Lynn A. Cooper

Personality Assessment

George Frank

Philosophical Views of Intelligence

Daniell N. Robinson

Philosophy for Children

Matthew Lipman

Piaget, Jean

*Chantal Bruchez-Hall**Howard E. Gruber*

Piagetian Theory of Development

*Bärbel Inhelder**Hermina Sinclair*

Practical Intelligence

Richard Wagner

Practice Effects

Alan S. Kaufman

Primary Mental Abilities Theory

John B. Carroll

Problem Finding

Mihaly Csikszentmihalyi

Problem Solving

Richard E. Mayer

Prodigies

David Henry Feldman

Profile Interpretation

Robert J. Gregory

Project Intelligence

Raymond S. Nickerson

Psychiatric Disorders

Robert A. Bornstein

Psychometric Theories of Intelligence

Nathan Brody

Psychometrics

Roger E. Millsap

Psychophysical Measures of Intelligence

Naftali Raz

Psychophysiological Measures of Intelligence

*John Kounios***Q**

Quick Measures of Intelligence

*Gary J. Robertson***R**

Race and Intelligence

*Edmund W. Gordon**Maitrayee Bhattacharyya*

Race and IQ Scores

Arthur R. Jensen

Radex Theory

Ruth Guttman

Rational Thinking

Jonathan Baron

Raven Progressive Matrices

John H. Court

Reaction Time

Philip Anthony Vernon

Reading

Charles A. Perfetti

Reasoning, Deductive

Ruth M. J. Byrne

Reasoning, Inductive

Richard E. Mayer

Reasoning, Moral

Elliot Turiel

Regional Differences in Intelligence

*Cecil R. Reynolds**Elizabeth Murdoch James*

Reliability

*Cecil R. Reynolds***S**

Savants

Michael J. A. Howe

Scholastic Assessment Tests

Sylvia T. Johnson

Schooling

Stephen J. Ceci

Serial and Parallel Processing

J. P. Das

Sex Chromosomal Abnormalities

John Money

Simon, Herbert A.

Kenneth Kotovsky

Situated Cognition

Gary M. Olson

Situated Intelligence

Martin E. Ford

Social Policy, Intelligence, and Testing

Michael J. Feuer

Socialization of Intelligence

Lynn Okagaki

Socioeconomic Status and Intelligence

Eric Turkheimer

Spatial Ability

David F. Lohman

Spearman, Charles Edward

Arthur R. Jensen

Speediness

Ted Nettelbeck

Stability of Intelligence

Bert Hayslip, Jr.

Stages of Cognitive Development

Charles J. Brainerd

Standardization

Malcolm Ree

Stanford-Binet Intelligence Scale, Fourth Edition

*Naomi G. Singer**Jerome M. Sattler*

Statistical Concepts

*Robert Cudeck**Lisa L. O'Dell*

Street Intelligence

Terezinha Nunes

Structure-of-Intellect Model

Talia Ben-Zeev

System of Multicultural Pluralistic Assessment

*Jane R. Mercer***T**

Terman, Lewis M.

John F. Feldhusen

Terman's Giftedness Study

David M. Fetterman

Test-Taking Strategies

Lloyd Bond

Testing in Government and Industry

Joseph Zeidner

Thomson, Godfrey W.

Lloyd G. Humphreys

Thomson's Random Overlap Theory

Lloyd G. Humphreys

Thorndike, Edward L.

Robert M. Thorndike

Thurstone, L. L.

Patricia Ruzgis

Triadic Theory of Ability Structure

Raymond B. Cattell

Triarchic Theory of Human Intelligence

Robert J. Sternberg

Twin Studies of Intelligence

Thomas J. Bouchard, Jr.

Two-Factor Theory

*Philip Anthony Vernon***U**

Underachievement

*Sandra Scarr***V**

Validity

Pamela Moss

Verbal Ability

Janet Starr Campito

Vernon, Philip E.

Philip Anthony Vernon

Vocational Abilities

*Jo-Ida C. Hansen**Deborah G. Betsworth*

Vygotskian Theories of Intelligence

*Herbert Pick**Julia B. Gippenreiter*

Vygotsky, Lev

Herbert Pick

Julia B. Gippenreiter

W

WAIS-R Subtests

James E. Lindemann

Wechsler, David

Allen Jack Edwards

Wechsler Scales of Intelligence

Aurelio Prifitera

Wisdom, Psychology of

Ursula Staudinger

Paul B. Baltes

Woodcock-Johnson Tests of Cognitive
Ability—Revised

Kevin S. McGrew

Work Force, Intelligence in the

Richard Wagner

Zack Loukides

Y

Yerkes, Robert M.

Richard T. von Mayrhauser

LIST OF CONTRIBUTORS

BRIAN P. ACKERMAN

University of Delaware

Motivation

PHILLIP L. ACKERMAN

University of Minnesota

Humphreys, Lloyd G.; Learning, Skill, and
Transfer

DAVID L. ALDERTON

Centers for Disease Control and Prevention, Atlanta, GA

Mechanical Ability

LOUISE BATES AMES

Gesell Institute of Human Development, New Haven, CT

Gesell, Arnold

WILLIAM H. ANGOFF

Educational Testing Service, Princeton, NJ

Brigham, Carl C.

DOROTHY M. ARAM

Emerson College

Language and Intelligence

LARS BÄCKMAN

Karolinska Institute, Stockholm

Compensatory Mechanisms

PAUL B. BALTES

*Max Planck Institute for Human Development and
Education, Berlin*

Aging and Intelligence; Wisdom, Psychology of

MARGARET J. BARCH

University of Louisville

Chromosomal Abnormalities

JONATHAN BARON

University of Pennsylvania

Rational Thinking

ALFRED A. BAUMEISTER

Vanderbilt University

American Association on Mental Retardation

JOHN M. BELMONT

University of Kansas

Mental Retardation, Organic

TALIA BEN-ZEEV

Yale University

Structure-of-Intellect Model

MICHELLE L. BENGSTON

Nova University

Dyscalculia

ARTHUR BENTON

Iowa City, IA

Brain, Pathologies of the; Goldstein, Kurt

JOHN W. BERRY

Queens University

Cross-cultural Variations in Intelligence

DEBORAH G. BETSWORTH

Counseling Center, Grand Forks, ND

Vocational Abilities

MAITRAYEE BHATTACHARYYA

Pomona, NY

Race and Intelligence

- ALLAN BLOOM
University of Louisville
Chromosomal Abnormalities
- LLOYD BOND
University of North Carolina, Greensboro
Test-taking Strategies
- INGWER BORG
Mannheim, Germany
Facet Theory
- MARC H. BORNSTEIN
Bethesda, MD
Infancy
- ROBERT A. BORNSTEIN
Ohio State University
Localization of Brain Function; Psychiatric Disorders
- DOUGLAS A. BORS
University of Toronto
Hebb's Theory of Intelligence
- THOMAS J. BOUCHARD, JR.
University of Minnesota
Twin Studies of Intelligence
- KENNETH S. BOWERS
University of Waterloo
Intuition
- CHARLES J. BRAINERD
Tucson, AZ
Child and Adolescent Intelligence; Children's Conservation Concepts; Stages of Cognitive Development
- NATHAN BRODY
West Hartford, CT
Psychometric Theories of Intelligence
- CHANTAL BRUCHEZ-HALL
New York, NY
Piaget, Jean
- GARY BUNCH
York University
Mainstreaming
- EARL C. BUTTERFIELD
University of Washington
Learning and Intelligence; Metacognition
- RUTH M. J. BYRNE
Trinity College (Dublin)
Reasoning, Deductive
- JANET STARR CAMPITO
Ballston Spa, NY
Verbal Ability
- JERRY S. CARLSON
University of California, Riverside
Dynamic Assessment of Mental Abilities
- JOHN B. CARROLL
Chapel Hill, NC
Primary Mental Abilities Theory
- RUFUS L. CARTER
Florida State University
Adulthood, Achievement in
- RAYMOND B. CATTELL
Honolulu, HI
Triadic Theory of Ability Structure
- STEPHEN J. CECI
Cornell University
Bioecological Theory of Intelligence; Schooling
- JEANNE S. CHALL
Harvard University
Illiteracy
- DONALD J. COHEN
Yale University
Autism
- LYNN A. COOPER
Columbia University
Perception
- JOHN H. COURT
Fuller Theological Seminary
Raven Progressive Matrices
- MIHALY CSIKSZENTMIHALYI
University of Chicago
Creativity; Problem Finding

- ROBERT CUDECK
University of Minnesota
Statistical Concepts
- MARY E. CURTIS
Father Flanagan's Boys' Home, Boys Town, NE
Illiteracy
- MARK H. DANIEL
American Guidance Service, Circle Pines, MN
Differential Ability Scales
- J. P. DAS
University of Alberta
Eastern Views of Intelligence; Luria, A. R.; Serial
and Parallel Processing
- JANET E. DAVIDSON
Lewis and Clark University
Insight
- LYLE DAVIDSON
New England Conservatory of Music
Musical Intelligence
- RENE V. DAWIS
University of Minnesota
Occupations
- PAUL DEBOECK
Catholic University, Louvain
Latent Trait Theory
- FRANK N. DEMPSTER
University of Nevada, Las Vegas
Interference
- FLORENCE DENMARK
New York, NY
Anastasi, Anne
- CURTIS K. DEUTSCH
Shriver Center, Waltham, MA
Down Syndrome
- DIANA DEUTSCH
University of California, San Diego
Musical Ability
- LISABETH F. DiLALLA
Southern Illinois University
Crime and Delinquency; Evolution of Human
Intelligence
- MARY DiSIBIO
California State University, Hayward
Adaptive Behavior, Assessment of
- ROGER A. DIXON
University of Victoria
Compensatory Mechanisms
- ALLEN JACK EDWARDS
Springfield, MO
Wechsler, David
- JULIE A. EISELE
Skidmore College
Language and Intelligence
- SUSAN EMBRETSON
University of Kansas
Latent Trait Theory
- RANDALL W. ENGLE
University of South Carolina
Memory
- ROBERT EPSTEIN
Cardiff by the Sea, CA
Behaviorism
- K. ANDERS ERICSSON
Florida State University
Memory, Exceptional
- HANS J. EYSENCK
Institute of Psychology, London
EEG Evoked Potentials
- JOHN F. FELDHUSEN
Purdue University
Terman, Lewis M.
- DAVID HENRY FELDMAN
Tufts University
Prodigies

- DAVID M. FETTERMAN
American Institutes for Research, Palo Alto, CA
Gifted and Talented Students, Evaluation of
Educational Programs for; Terman's
Giftedness Study
- MICHAEL J. FEUER
Washington, DC
Social Policy, Intelligence, and Testing
- REUVEN FEUERSTEIN
Hadassah Wizo Canada Research Institute, Jerusalem
Learning Potential Assessment Device
- RONALD A. FINKE
Texas A&M University
Imagery
- JAMES R. FLYNN
University of Otago, Dunedin
IQ Gains over Time
- MARTIN E. FORD
George Mason University
Social Intelligence
- ANTONIA A. FORSTER
Oregon Health Sciences University
Learning Disability
- GEORGE FRANK
New York, NY
Personality Assessment from Tests of Intelligence
- JOHN FREMER
Washington Crossing, PA
Group Tests
- JENNIFER J. FREYD
University of Oregon
Imagery
- HOWARD S. FRIEDMAN
University of California, Riverside
Health and Intelligence
- HOWARD GARDNER
Harvard University
Multiple Intelligences
- SUSAN A. GELMAN
University of Michigan
Competence Versus Performance
- KENNETH J. GILHOOLY
University of Aberdeen
Knowledge
- JULIA B. GIPPENREITER
Moscow State University, Moscow
Vygotskian Theories of Intelligence; Vygotsky, Lev
- CLAIRE GOLOMB
University of Massachusetts, Boston
Artistic Ability: Children's Drawings
- EDMUND W. GORDON
Gordon & Gordon Associates, Pomona, NY
Race and Intelligence
- ROBERT J. GREGORY
University of Idaho
Classification of Intelligence; Profile Interpretation
- HOWARD E. GRUBER
Teachers College, Columbia University
Piaget, Jean
- JAN-ERIC GUSTAFSSON
University of Goteborg
General Intelligence; Hierarchical Theories of
Intelligence
- TERRY B. GUTKIN
University of Nebraska
Asian Americans; Hispanics; Japanese
- RUTH GUTTMAN
Hebrew University
Guttman, Louis E.; Radex Theory
- JAMES S. GYURKE
The Psychological Corporation, San Antonio, TX
McCarthy Scales of Children's Abilities
- KENJI HAKUTA
Stanford University
Bilingualism

- DIANE F. HALPERN
California State University, San Bernardino
Gender Differences in Intellectual Abilities
- KERRY DES. HAMSHER
Neuropsychology Clinic, Milwaukee, WI
Benton, Arthur L.
- JO-IDA C. HANSEN
University of Minnesota
Vocational Abilities
- ANN E. HARMAN
University of North Carolina, Greensboro
Test-taking Strategies
- PATTI L. HARRISON
University of Alabama
Draw-A-Figure Test
- BERT HAYSLIP, JR.
University of North Texas
Age and the Content of Intelligence Tests; Stability of Intelligence
- LOUIS M. HERMAN
University of Hawaii
Animal Intelligence: Historical Perspectives and Contemporary Approaches
- JOSEPH H. HERSH
University of Louisville
Chromosomal Abnormalities
- CHRISTOPHER HERTZOG
Georgia Institute of Technology
Cohort Effects
- ROBERT M. HODAPP
University of California, Los Angeles
Mental Retardation, Cultural-Familial
- PAUL J. HOFER
Federal Judicial Center, Washington, DC
Legal Issues in Intelligence
- ALICE STERLING HONIG
Syracuse University
Intervention, Infant and Preschool
- JOHN L. HORN
University of Southern California
Fluid and Crystallized Intelligence, Theory of
- FRANCES DEGEN HOROWITZ
City University Graduate Center, New York, NY
Giftedness
- MICHAEL J. A. HOWE
University of Exeter
Genius; Savants
- LLOYD G. HUMPHREYS
University of Illinois
Measurement and Prediction of Intelligence; Thomson, Godfrey, H.; Thomson's Random Overlap Theory
- JOHN E. HUNTER
Lansing, MI
General Aptitude Test Battery
- GEORGE W. HYND
University of Georgia
Aphasia
- BÄRBEL INHELDER
University of Geneva
Piagetian Theory of Intellectual Development
- CARROLL E. IZARD
University of Delaware
Motivation
- ELIZABETH MURDOCH JAMES
Texas A&M University
Regional Differences in Intelligence
- ARTHUR R. JENSEN
University of California, Berkeley
Eysenck, Hans J.; Galton, Francis; Race and IQ Scores; Spearman, Charles E.
- SYLVIA T. JOHNSON
Howard University
Scholastic Assessment Tests
- LYLE V. JONES
University of North Carolina, Chapel Hill
Carroll, John B.

- ALAN S. KAUFMAN
Escondido, CA
Case Reports; Factor Analysis of the Wechsler and
Stanford-Binet Scales; Practice Effects
- NADEEN L. KAUFMAN
Escondido, CA
Case Reports
- DANIEL P. KEATING
Center for Applied Cognitive Science, Toronto, Ontario
Contextualist Theories of Intelligence
- PAUL KLINE
University of Exeter
Cattell, Raymond B.
- NATHAN KOGAN
New School for Social Research
Cognitive Styles
- KENNETH KOTOVSKY
Carnegie-Mellon University
Simon, Herbert A.
- JOHN KOUNIOS
Tufts University
Psychophysiological Measures of Intelligence
- PATRICK C. KYLLONEN
Brooks Air Force Base, TX
Abstraction; Information Processing
- GISELA LABOUVIE-VIEF
Wayne State University
Dialectical Thinking
- DANIEL M. LANDERS
Arizona State University
Athletic Ability
- GERALD LARSON
Navy Personnel R&D Center, San Diego, CA
Armed Services Vocational Aptitude Battery
- INGRID N. LECKLITER
Oregon Health Sciences University
Dyslexia
- HENRY LELAND
Ohio State University
Adaptive Behavior; Goddard, Henry H.
- JAMES E. LINDEMANN
Oregon Health Sciences University
WAIS-R Subtests
- ULMAN LINDENBERGER
*Max Planck Institute for Human Development and
Education, Berlin*
Aging and Intelligence
- MATTHEW LIPMAN
Montclair State College
Philosophy for Children
- LEWIS P. LIPSITT
Brown University
Bayley, Nancy; Motor Ability
- JOHN C. LOEHLIN
University of Texas at Austin
Genetics, Behavior
- DAVID F. LOHMAN
University of Iowa
Spatial Ability
- ZACK LOUKIDES
Florida State University
Work Force, Intelligence in the
- COLIN M. MACLEOD
University of Toronto
Hebb, Donald O.
- JOAN C. MARTIN
University of Washington
Birth Defects
- JOSEPH D. MATARAZZO
Oregon Health Sciences University
Binet, Alfred; Biological Measures of Intelligence;
Deviation Quotient; Intelligence Quotient;
Mental Age; Mental Measurements Yearbook
- JAMES MAXEY
American College Test, Iowa City, IA
ACT Assessment

LIST OF CONTRIBUTORS

- RICHARD E. MAYER
University of California, Santa Barbara
Mathematical Ability; Problem Solving; Reasoning,
Inductive
- JOHN J. MCARDLE
University of Virginia
Factor Analysis
- KEVIN S. MCGREW
Saint Cloud State University
Woodcock-Johnson Tests of Cognitive Ability-
Revised
- FREDERICK L. MCGUIRE
Orange, CA
Army Alpha and Beta Tests of Intelligence
- MANFRED MAIER
Minneapolis, MN
Broca, Paul
- DAVID KENT MERCER
Minneapolis, MN
Broca, Paul
- JANE R. MERCER
University of California, Riverside
System of Multicultural Pluralistic Assessment
- PHILIP R. MERRIFIELD
New York, NY
Guilford, J. P.
- ROGER E. MILLSAP
Brooklyn, NY
Psychometrics
- WILEY MITTENBERG
Nova University
Dyscalculia
- JOHN MONEY
Baltimore, MD
Sex Chromosomal Abnormalities
- PAMELA A. MOSS
University of Michigan
Validity
- BETH MURDOCK
Texas A&M University
Mercer, Jane R.
- TED NETTELBECK
University of Adelaide
Speediness
- RAYMOND S. NICKERSON
Bedford, MA
Interventions, Later; Project Intelligence
- LORRIE A. NIELSON
Southern Illinois University
Crime and Delinquency
- SARA JO NIXON
Center for Alcohol Studies, Oklahoma City, OK
Alcohol and Alcohol Abuse
- TEREZINHA NUNES
University of London
Street Intelligence
- LISA L. O'DELL
University of Minnesota
Statistical Concepts
- JOHN U. OGBU
University of California, Berkeley
Culture and Intelligence
- JAMES R. P. OGLOFF
Simon Fraser University
Criminality
- LYNN OKAGAKI
Purdue University
Socialization of Intelligence
- GARY M. OLSON
University of Michigan
Situated Cognition
- HEATHER CARMICHAEL OLSON
University of Washington
Fetal Alcohol Syndrome
- ADAM A. PACK
University of Hawaii
Animal Intelligence: Historical Perspectives and
Contemporary Approaches

- OSCAR A. PARSONS
Center for Alcohol Studies, Oklahoma City
Alcohol and Alcohol Abuse
- CAROL PATRICK
Southern Illinois University
Evolution of Human Intelligence
- CHARLES A. PERFETTI
University of Pittsburgh
Reading
- HERBERT L. PICK, JR.
University of Minnesota
Vygotskian Theories of Intelligence; Vygotsky, Lev
- ROBERT PLOMIN
Penn State University
Nature, Nurture, and Development
- ANTHONY M. PODRAZA
Maine Head Trauma Center, Bangor
Localization of Brain Function
- AURELIO PRIFITERA
San Antonio, TX
Wechsler Scales of Intelligence
- MITCHELL RABINOWITZ
Fordham University
Connectionism
- CRAIG T. RAMEY
University of Alabama
Abecedarian Project
- NAFTALI RAZ
Memphis State University
Psychophysical Measures of Intelligence
- MALCOLM REE
Brooks Air Force Base, TX
Job Performance; Standardization
- RALPH M. REITAN
Reitan Neuropsychology Laboratory, South Tucson, AZ
Halstead, Ward Campbell; Neuropsychology,
Clinical
- LAUREN RESNICK
University of Pittsburgh
Robert Glaser
- CECIL R. REYNOLDS
Texas A&M University
Bias in Testing; Jensen, Arthur R.; Kaufman
Assessment Battery for Children; Mercer,
Jane R.; Regional Difference in
Intelligence; Reliability
- HENRY N. RICCIUTI
Cornell University
Infant Tests as Measures of Early Competence;
Nutrition
- DAVID RINDSKOPF
Chestnut Ridge, NY
Experimental Design
- GARY J. ROBERTSON
Circle Pines, MN
Quick Measures of Intelligence
- DANIEL N. ROBINSON
Georgetown University
Philosophical Views of Intelligence
- JOSEPH LEE RODGERS
University of Oklahoma
Birth Order, Spacing, and Family Size; Family
Environments
- KAREN B. ROGERS
University of St. Thomas
Ability Grouping
- DAVID C. ROWE
University of Arizona
Birth Order, Spacing, and Family Size; Family
Environments
- SCOTT S. RUBIN
University of Georgia
Aphasia
- DUANE M. RUMBAUGH
Georgia State University
Animal Intelligence: Primate
- PATRICIA RUZGIS
Yale University
Thurstone, L. L.

LIST OF CONTRIBUTORS

- JEROME M. SATTLER
San Diego State University
Individual Tests; Stanford-Binet Intelligence Scale,
Fourth Edition
- SANDRA SCARR
University of Virginia
Burt, Cyril L.; Culture-Fair and Culture-Free
Tests; Underachievement
- DAVID G. SCHAEFER
Simon Fraser University
Criminality
- ROGER SCHANK
Northwestern University
Artificial Intelligence
- HANNELORE SCHOCK
Birmingham, AL
Draw-A-Figure Test
- PETER SCHÖNEMANN
West Lafayette, IN
Heritability
- CAROL J. SCHRAMKE
Pittsburgh, PA
Agnosia
- HARRIET SHAKLEE
Seattle, WA
Decision Making and Judgment
- ARTHUR P. SHIMAMURA
University of California, Berkeley
Brain
- JOAN W. SHIMAMURA
California Pacific Medical Center, San Francisco
Brain
- ROBERT S. SIEGLER
Carnegie-Mellon University
Development, Cognitive
- HERMINA SINCLAIR
Archives Jean Piaget, University of Geneva
Piaget's Theory of Intellectual Development
- NAOMI G. SINGER
San Diego State University
Individual Tests; Stanford-Binet Intelligence Scale,
Fourth Edition
- EDWARD J. SLAWSKI
The Psychological Corporation, San Antonio, TX
Error of Measurement
- RICHARD E. SNOW
Stanford University
Abilities and Aptitudes; Aptitude-Treatment
Interaction; Cronbach, Lee J.
- ROBERT J. SOLSO
University of Nevada
Cattell, James McKeen
- CHARLES D. SPIELBERGER
University of South Florida
Anxiety
- LAZAR STANKOV
University of Sydney
Auditory Abilities
- JULIAN C. STANLEY
Johns Hopkins University
Leta S. Hollingworth
- URSULA STAUDINGER
*Max Planck Institute for Human Development and
Education, Berlin*
Wisdom, Psychology of
- ROBERT J. STERNBERG
Yale University
Triarchic Theory of Human Intelligence
- PETER SUEDFELD
University of British Columbia
Complexity, Cognitive
- LISA A. SUZUKI
University of Oregon
Asian Americans; Ethnicity, Race, and Measured
Intelligence; Hispanics; Japanese
- SUMNER J. SYDEMAN
University of South Florida
Anxiety
-

- ROBERT M. THORNDIKE
Western Washington University
Thorndike, Edward L.
- WILLIAM C. TIRRE
Brooks Air Force Base, TX
Bond Sampling Theory of Human Abilities
- BRUCE TORFF
Boston College
Musical Intelligence
- ELLIOT TURIEL
University of California, Berkeley
Reasoning, Moral
- ERIC TURKHEIMER
University of Virginia
Socioeconomic Status and Intelligence
- INA Č. UŽGIRIS
Clark University
Hunt, Joseph McVicker
- JAAN VALSINER
University of North Carolina
Interactionist Views on Intelligence
- PHILIP ANTHONY VERNON
University of Western Ontario
Reaction Time; Two-Factor Theory;
Vernon, Philip E.
- FRED R. VOLKMAR
Yale University
Autism
- RICHARD T. VON MAYRHAUSER
Harvard University
Doll, Edgar A.; Yerkes, Robert M.
- DAMIAN A. VRANIAK
University of Wisconsin, Madison
Ethnicity, Race, and Measured Intelligence;
Native Americans
- RICHARD WAGNER
Florida State University
Practical Intelligence; Work Force, Intelligence
in the
- MARK F. WARD
V.A. Medical Center, Portland, OR
Attention Deficit Hyperactivity Disorder;
Hyperactivity
- DAVID A. WASHBURN
Georgia State University
Animal Intelligence: Primate
- GARY L. WENK
University of Arizona
Drugs and Intelligence
- DEAN K. WHITLA
Harvard University
College Board
- WENDY M. WILLIAMS
Yale University
Parenting and Intelligence
- MELVIN N. WILSON
University of Virginia
African Americans
- DEBORAH WOLFSON
Reitan Neuropsychology Laboratory, South Tucson, AZ
Neuropsychology, Clinical
- RICHARD W. WOODCOCK
Charlottesville, VA
Fluid and Crystallized Theory of Intelligence,
Measures of; Norms
- PAMELA WOODLEY-ZANTHOS
Vanderbilt University
American Association on Mental Retardation
- STEVEN YANTIS
Johns Hopkins University
Attention
- JOSEPH ZEIDNER
George Washington University
Testing in Government and Industry
- MICHAEL ZIEKY
Educational Testing Service, Princeton, NJ
Achievement Testing

A

ABECEDARIAN PROJECT The Abecedarian Project was developed by Craig Ramey and his colleagues at the University of North Carolina at Chapel Hill, beginning in 1972, to better understand the modifiability of early intellectual development as a function of various early childhood experiences. The main focus of the project is intervention and enrichment for the prevention of intellectual decline into the borderline and retarded categories, which are frequently associated with very poor family economic and social circumstances.

The Abecedarian Project has used a randomized, controlled experimental design to address three primary specific aims: 1. to determine whether developmental retardation and school failure can be prevented in children from socially and economically high-risk families by means of family supports, including high-quality educational day care beginning in early infancy; 2. to determine whether a follow-through program for kindergarten and early elementary school is necessary and sufficient to maintain preschool intellectual gains in high-risk children; and 3. to determine whether intervention at kindergarten and early elementary-school age can alone significantly improve academic and/or intellectual performance in children who did not have intensive preschool intervention.

Eligible children ($N = 111$) were identified typically during the last trimester (three months) of pregnancy,

using a multifactor high-risk index (Ramey & Smith, 1977), and randomly assigned shortly after birth to one of two experimental conditions. In the preschool control condition, children received developmental and pediatric surveillance and referral, if necessary, and nutritional supplements; their families received family support services from social workers. In the preschool early-intervention condition, these same services were provided, and the children received a systematic educational day-care program beginning at 6 weeks of age and lasting until they were enrolled in public school kindergarten. Just before kindergarten entry, the early-intervention and control groups were each randomly divided into two groups, and one half of each original group received a special home-school resource program for the first three years of public school. The research design is depicted in Figure 1.

Key factors of the preschool intervention included low teacher-child ratios; a specific educational program on a full-day, year-round basis; daily transportation; good nutrition; and medical surveillance. Key features of the home-school resource program included individualized educational plans developed by master teachers, regular consultation with parents and classroom teachers to improve mutual understanding of educational goals and specific teaching methods, and individualized summer programs, including academic and summer camp experiences.

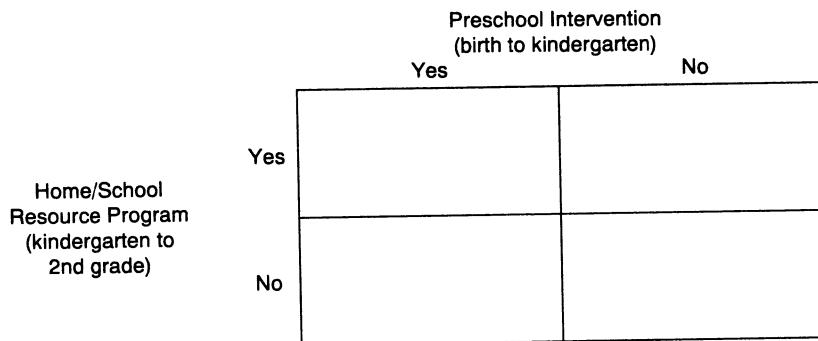


Figure 1
Research design of the Abecedarian Project

Results from developmental follow-up during the preschool period (Ramey & Campbell, 1984; Ramey, Yeates, & Short, 1984) indicated that the intensive preschool early intervention significantly prevented intellectual declines into mental subnormality. At age 6 months through 54 months, the intelligence quotients (IQs) of early-intervention children ranged from 7.9 to 20.1 points higher than those of control children when the effects of maternal mental retardation and home environment were statistically controlled; at every age, a greater proportion of the experimental-program children scored in the normal range for IQ (i.e., above 84). In 13 children with retarded mothers, not one of the 6 experimental-program children, but 6 of 7 control children, had IQ scores below normal (Martin, Ramey, & Ramey, 1990).

Follow-up at age 8, at the end of the school-age phase of the experiment, revealed continued positive effects of preschool intervention on intellectual performance as measured by IQ tests as well as by standardized tests of academic achievement. In addition, the likelihood of failing a grade because of academic unreadiness for grade promotion was reduced (Horacek et al., 1987; Ramey & Campbell, 1992). The effects of school-age intervention were weaker than preschool intervention on academic achievement and non-existent for IQ.

Follow-up at age 12 (Ramey, 1993; Ramey et al., 1992), four years after all educational intervention had been terminated, showed that the Abecedarian early-intervention group continues to score near the national average and significantly higher (5.1 IQ points on the WISC-R) than do their randomized controls.

The findings at age 12 also indicate the preschool-intervention children had higher achievement scores in reading and mathematic achievement, a 49 percent reduction in retention in grade (i.e., grade failure) and a 71 percent reduction in IQs of 85 or less. The minimal effects of the school-age intervention on academic achievements at age 8 were no longer detectable at age 12, thus buttressing the hypothesis that early intervention for high-risk children is needed to improve their developmental outcomes with respect to IQ and academic achievement.

(See also: INTERVENTION, INFANT AND PRESCHOOL.)

BIBLIOGRAPHY

- HORACEK, H. J., RAMEY, C. T., CAMPBELL, F. A., HOFFMAN, K. P., & FLETCHER, R. H. (1987). Predicting school failure and assessing early interventions with high-risk children. *Journal of the American Academy of Child Psychiatry*, 26(5), 758-763.
- MARTIN, S. L., RAMEY, C. T., & RAMEY, S. L. (1990). The prevention of intellectual impairment in children with impoverished families: Findings of a randomized trial of educational daycare. *American Journal of Public Health*, 80, 844-847.
- RAMEY, C. T. (1993). A rejoinder to Spitz's critique of the Abecedarian experiment. *Intelligence*, 17, 25-30.
- RAMEY, C. T., & CAMPBELL, F. A. (1984). Preventive education for high-risk children: Cognitive consequences of the Carolina Abecedarian Project. *American Journal of Mental Deficiency*, 88(5), 515-523.

- RAMEY, C. T., & CAMPBELL, F. A. (1992). Poverty, early childhood education, and academic competence: The Abecedarian experiment. In A. Huston (Ed.), *Children in poverty* (pp. 190–221). New York: Cambridge University Press.
- RAMEY, C. T., CAMPBELL, F. A., BRYANT, D. M., BURCHINAL, M., SPARLING, J., & WASIK, B. H. (1992). Early intervention and long-term predictors of school status. In F. Parker et al. (Eds.), *New directions in child and family research* (pp. 170–174). Washington, DC: Administration on Children, Youth, and Families.
- RAMEY, C. T., & SMITH, B. J. (1977). Assessing the intellectual consequences of early intervention with high-risk infants. *American Journal of Mental Deficiency, 8*(4), 318–324.
- RAMEY, C. T., YEATES, K. O., & SHORT, E. J. (1984). The plasticity of intellectual development: Insights from preventive intervention. *Child Development, 55*, 1913–1925.
- YEARGAN-ALLSOPP, M., MURPHY, C. C., OAKLEY, G. P., & SIKES, R. K. (1992). A multiple source method for studying the prevalence of developmental disabilities in children: The metropolitan Atlanta developmental disabilities study. *Pediatrics, 89*, 624–630. (Published errata appears in *Pediatrics, 1992, 90*, 1001.)

CRAIG T. RAMEY

ABILITIES See ARTISTIC ABILITY; ATHLETIC ABILITY; AUDITORY ABILITIES; MATHEMATICAL ABILITY; MECHANICAL ABILITY; MOTOR ABILITY; MUSICAL ABILITY; SPATIAL ABILITY; VERBAL ABILITY; VOCATIONAL ABILITIES.

ABILITIES AND APTITUDES This entry defines human abilities in relation to intelligence, discusses the concept of aptitude, and summarizes the evidence on ability differences as predictors of learning and performance in education and work. It does not treat each kind of specialized ability in detail.

ABILITIES AND INTELLIGENCE

An ability is a power to perform some specified act or task, either physical or mental. Different tasks call for somewhat different abilities, but there are broad abilities that apply to many kinds of performance tasks and narrow abilities that relate only to one specialized kind of task. Throughout most of the twentieth century, a major line of research in differential psychology

has used the methods of factor analysis to identify the arrays of mental abilities thought to be constituents of general intelligence. J. B. CARROLL'S (1993) massive review and reanalysis of much of this evidence has resulted in an integrated taxonomy of these abilities. His hierarchical model lists a large number of specialized first-order abilities, which belong to eight second-order ability domains, called *fluid intelligence*, *crystallized intelligence*, *general memory and learning*, *broad visual perception*, *broad auditory perception*, *broad retrieval ability*, *broad cognitive speediness*, and *processing speed*. These domains connect in turn to the general intelligence construct at the third level. A similar hierarchical model and list of constituent abilities come from J. L. Horn's (1989) extensive work. Furthermore, J. E. Gustafsson (1988, 1989) has demonstrated that fluid inductive reasoning ability can equate with GENERAL INTELLIGENCE, that controlling statistically for the relation of general intelligence to crystallized and visual perception abilities then reproduces P. E. Vernon's (1950) earlier hierarchical model, and that both the broad higher-order and the narrow lower-order abilities need to be studied together in analyzing individual differences in learning and performance (see also FLUID AND CRYSTALLIZED INTELLIGENCE, THEORY OF). In other words, evidence shows that the hierarchical model of intellectual abilities is theoretically coherent, empirically replicable, and practically useful. Research with methods other than factor analysis gives a consistent picture (Snow, Kyllonen, & Marshalek, 1984). Thus, specific intellectual ability constructs and measures may come from different levels of hierarchical taxonomy for use in further research as well as in practical settings where ability differences among persons are a concern. A taxonomy of physical and psychomotor abilities based on factor analytic research also exists; specific abilities involved in strength, coordination, dexterity, attention, and control are defined, and measures are exemplified (see Fleishman & Quaintance, 1984).

ABILITIES AS APTITUDES AND ACHIEVEMENTS

Although the terms *ability* and *aptitude* often occur as synonyms, the concept of aptitude includes any enduring personal characteristics that are propaedeutic

to successful performance in some particular situation. This definition includes affective, conative, and personality characteristics as well as cognitive and psychomotor abilities (Snow, 1992). However, cognitive abilities are a particularly important source of aptitude for learning and performance in many school and work situations.

Personnel decisions in industry, government, and the military often rely on specific ability measures as indicators of aptitude. The measures are validated for particular uses by showing that they predict important criteria of job success. Persons are then selected for employment, training, or promotion in part on the basis of their ability test scores. Such scores also serve classification decisions, as when different persons are assigned to different jobs according to their ability profiles (see JOB PERFORMANCE; WORK FORCE, INTELLIGENCE IN THE).

Ability measures represent aptitudes for learning in education and training settings when they predict important criteria of achievement. Again, the measures can be used to select applicants predicted to be most successful or to classify students into different instructional treatments designed to be adaptive to particular ability types or levels. Ability measures are also used to evaluate the suitability of particular educational programs for different persons. Some are designed for diagnosis of specific learning disabilities.

In education, however, and also in some training situations, many of these same ability constructs are targets of instruction. The instructional goal is to develop specific abilities for use in later learning and performance situations. Reading and mathematical abilities are the obvious examples; they represent goals much emphasized throughout primary and secondary education. Special programs address other kinds of abilities as well. Thus, abilities are achievements from past learning just as they are at the same time aptitudes for future learning. Education is in this sense an aptitude development program, and ability measures can serve to index both its input and its output functions. Moreover, different kinds of educational programs may both attract and help develop broadly different aptitude profiles. There is some evidence that more classical academic programs develop stronger verbal-crystallized abilities, whereas more technical programs develop stronger spatial-mechanical abilities

(Balke-Aurell, 1982). Most abilities develop from extensive experience across learning history, not from specific courses. And ability differences are influenced by genetic factors as well as by experience.

GENERAL VERSUS SPECIAL ABILITIES

Controversy exists over the relative importance of general versus special abilities in predicting academic and job success. Some researchers argue that only general ability or a very few broad abilities are needed to account for individual differences in performance in a wide variety of jobs and instructional and training programs. A further conclusion derived from this view is that ability measures used for selection or classification purposes need not be validated in each application because the validity of more general measures generalizes across many jobs and settings, as well as across time. A more cautious view argues that the relevance of more specialized abilities for particular performance situations is often underestimated because of the effects of prior selection processes including self-selection on the ability distributions being studied, and the use of inappropriate statistical procedures. The number of abilities to be investigated and the need for local validation of their measures are important theoretical questions that also have major practical importance. Their answers influence both the costs and the legalities involved in educational and employment selection procedures.

Special ability requirements differ across job families, educational programs, and performance settings, even though general ability measures are the strongest predictors of performance in many complex jobs and in many educational programs. Both kinds of measures must be present in a study to evaluate properly the importance of either, as Gustafsson (1989) demonstrated. Perhaps the most important uses of special ability differences ultimately will be in classification of persons between available jobs or programs rather than in selection for one, but this possibility has been least studied. The debate, its many complexities, and the needs for further research are well summarized by L. J. CRONBACH (1990).

(See also: ADAPTIVE BEHAVIOR.)

BIBLIOGRAPHY

- BALKE-AURELL, G. (1982). *Changes in ability as related to educational and occupational experience*. Göteborg, Sweden: Acta Universitatis Gothoburgensis.
- CARROLL, J. B. (1993). *Human cognitive abilities*. New York: Cambridge University Press.
- CRONBACH, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- FLEISHMAN, E. A., & QUAINANCE, M. K. (1984). *Taxonomies of human performances: The description of human tasks*. Orlando, FL: Academic Press.
- GUSTAFSSON, J. E. (1988). Hierarchical models of the structure of cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 35–71). Hillsdale, NJ: Erlbaum.
- GUSTAFSSON, J. E. (1989). Broad and narrow abilities in research on learning and instruction. In R. Kanfer, P. L. Ackerman, & R. Cudeck (Eds.), *Abilities, motivation, and methodology* (pp. 203–237). Hillsdale, NJ: Erlbaum.
- HORN, J. L. (1989). Cognitive diversity: A framework of learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research*, New York: W. H. Freeman.
- SNOW, R. E. (1992). Aptitude theory: Yesterday, today and tomorrow. *Educational Psychologist*, 27, 5–32.
- SNOW, R. E., KYLLONEN, P. C., MARSHALEK, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 2, pp. 47–104). Hillsdale, NJ: Erlbaum.
- VERNON, P. E. (1950). *The structure of human abilities*. London: Methuen.

RICHARD E. SNOW

ABILITY GROUPING Grouping by ability has been documented as a practice in U.S. schools since 1919. At that time, educators in Detroit used the achievement test scores of all children in this large city school district to place them in X, Y, or Z tracks. The top and bottom 20 percent of the students were placed in the X and Z tracks, respectively, while the remaining 60 percent were in the Y track. All three tracks, however, studied the same curriculum. The Detroit model was widely used throughout the United States until about 1970. Studying the effects of ability grouping through controlled experiments may have

first occurred in 1927 in Salt Lake City. Students were assigned to homogeneous ability groups or the control (mixed-ability classes), and the effects on achievement were measured after one year. For this study, the ability-grouped students outperformed their controls by two grade-equivalent school months.

The issues surrounding ability grouping proved controversial in most of the decades in which it has been practiced. In the 1930s, P. A. Witty and L. W. Wilkins (1933) reviewed extant literature and concluded that ability grouping was more helpful than harmful, although there might be more effective strategies for boosting the achievement of gifted students. By the 1960s, the body of actual research, not to mention the proliferation of articles in the popular and professional press, led A. H. Passow (1962) to write a classic article, "The Maze of the Research on Ability Grouping." In it, he concluded that the body of literature was at that point almost too difficult to put together in any reasoned, scholarly way because of inherent differences in study design and scope, sample characteristics and size, treatment implementation, measures used, subject areas studied, and forms of grouping by ability selected for each study.

Concerns have been raised in the past decade once again about the efficacy of grouping students by their levels of ability. In fact, considerably different conclusions are reached in the research journals and the professional journals, which are aimed at school practitioners and decision makers. The controversy has been exacerbated by the national movements for school reform, concerns for social equity, a focus on the affective issues in the socialization of school children, and concerns for raising the performance and expectation levels of at-risk children.

The remainder of this paper attempts to clarify why there have been perceived contradictions in educators' understanding of the benefits and limitations of ability grouping. More important, it attempts to address a variety of issues about this practice, which must be understood before conclusions can be drawn. Grouping issues cluster around five general themes: (1) research versus professional literature; (2) tracking versus ability grouping; (3) ability/intelligence versus achievement grouping; (4) grouping for management versus curriculum facilitation; and (5) full versus part-time grouping arrangements.

**THE RESEARCH VERSUS
PROFESSIONAL ISSUE: WHICH BODY
OF LITERATURE IS RIGHT?**

The ability grouping literature approaches 750 works of various forms. Of this number, approximately 100 are quantitative reports of research, another 75 are qualitative analyses of research, and the remainder represent opinion, program descriptions, and persuasive essays based on personal experiences, pro or con. Attempts to interpret the research body have been fairly consistent when they have clustered studies asking similar research questions about grouping. Attempts of earlier decades to cluster research findings relied on reviews of research, but by the time Passow expressed frustration about this issue, the numbers of contradictory studies were so large that representative reviews or even box-score counts of positive and negative research conclusions were inadequate to interpret these large bodies of research fairly.

Educators, who are confronted with the huge body of literature on ability grouping, ultimately resort to readable syntheses or interpretations of it. Unfortunately, the syntheses published in reputable research and research review journals have drawn one set of conclusions, and the syntheses prepared for professional journals subscribed to by school principals and classroom teachers have drawn another set of conclusions. Part of this difference in conclusions can be blamed on the result of simplifying first-hand information for a relatively untrained audience of readers; such simplifications may inadvertently remove some element of truth and distort the overall findings. This may occur even when the same author attempts to simplify his own study results. For example, the following two sets of conclusions *by the same author* were published within the same year—the first, the actual conclusions drawn in his research study report for *Review of Educational Research* and the second, a simplification of that study for a professional journal:

Evidence from 17 comparisons in 13 matched equivalent and 1 randomized study clearly indicates that assigning students to self-contained classes according to general achievement or ability does not enhance student achievement in the elementary school. (Slavin, 1987, p. 328)

The achievement effects of ability-grouped class assignments are essentially zero at the elementary level and are very slight at the secondary level. There is some evidence that high achievers may gain from ability grouping at the expense of low achievers. (Slavin, 1987, p. 34)

One can only imagine the simplification effect as a game of telephone, in which a message passed from one teller to the next becomes increasingly confused by each teller's attempts to make meaning of the confusion to pass the message on to the next teller. In the body of professional literature this appears to have taken place from article to article. Before long, the truth may be fully obscured in the retelling.

In 1976, Glass developed the research synthesis procedure he termed meta-analysis, whereby large groups of research studies would be reduced to a common metric, called effect size. With slight adjustments to this procedure (weighting for sample-size differences, stricter inclusion criteria, and testing for homogeneity among studies of poorer quality) the synthesis procedure has improved the potential to make sense of large bodies of research. As individual studies are reduced to this common metric, they can be averaged to produce an estimate of general effect. For achievement, for example, an effect size of .33, considered a small to moderate gain, can be interpreted as a proportion of the test's standard deviation unit. Since most test batteries were designed with an expected standard deviation unit of one for each school year, an effect size of .33 would suggest that an additional one-third of a grade-equivalent school year would have been produced by ability grouping, compared to heterogeneous grouping (Glass, 1976; Cohen, 1977; Wolf, 1986).

**THE TRACKING VERSUS
ABILITY-GROUPING ISSUE: WHAT
DO WE MEAN WHEN WE SAY
ABILITY GROUPING?**

Much of the misunderstanding about ability grouping coming from the current spate of literature on the subject can be attributed to the use of some terms as synonyms. In *Turning Points: Preparing Youth for the 21st Century* (published by the Carnegie Task Force on the

Education of Young Adolescents) (Hornbeck, 1987), in the work of Paul George (1988), and the works of Jeannie Oakes (1985), care was taken at the beginning of their treatises to define the words used to describe grouping, but as the works progressed, the terms *tracking* and *ability grouping* became synonymous. Slavin (1987; 1990) and the Kuliks (1982; 1984; 1985; 1990; 1992) have clarified the differential effects of a variety of grouping-by-ability strategies. For tracking (the full-time placement of students into a three-track system, whether called vocational, regular, and college prep or low, average, and high tracks), the effect sizes appear to hover around zero (with differential effects reported for gifted children in such a full-time tracked arrangement). For regrouping for specific instruction (whereby students are placed in specific high, average, or low classes subject by subject or are placed in courses with different names, subject by subject; e.g., high school students when enrolling in chemistry may be placed in Kitchen Chemistry, General Chemistry, or Theoretical Chemistry), the reported effect sizes range from zero in Slavin's two best-evidence syntheses in some academic subjects to .34 in other subjects, such as mathematics. For cross-grade grouping (whereby students study a common curriculum according to their individual progress in that subject rather than according to grade-level expectations), the mean effect sizes range from 26 percent across all grade levels (Kulik & Kulik, 1990; 1992) to .45 at the elementary level, but zero at the secondary level in Slavin's (1987; 1990) work. For within-class ability grouping (whereby the classroom teacher sorts the ability levels within one classroom into labeled groups such as the Bluebirds, Eagles and Grackles, for smaller group instruction in specific subjects), mean effect sizes have ranged from zero in reading and .34 in mathematics (Slavin, 1987), to .62 across all academic areas for gifted students clustered as a small group within the classroom (Kulik & Kulik, 1990). For gifted students too, Vaughn, Feldhusen, and Asher's (1991) meta-analysis on pullout-enrichment grouping reported mean effect sizes of .65 for achievement, .44 for critical thinking, and .32 for creativity gains over intellectual peers who remained in the mainstreamed classroom. The differences in effects from one form of grouping by ability to another are great.

THE ABILITY/INTELLIGENCE VERSUS ACHIEVEMENT ISSUE: SHOULD WE GROUP BY INTELLIGENCE LEVEL OR BY ACHIEVEMENT LEVEL?

To answer this question of grouping by intelligence or grouping by achievement level, one can look at the meta-analyses and best-evidence syntheses produced in the 1980s for some answers. The manner in which studies have been included or excluded/separated for these syntheses helps to explain some differences in conclusions. In Slavin's best-evidence synthesis of the controlled studies on grouping by achievement level (in other words, children grouped by their test scores on standardized or local-subject-matter tests of achievement), the benefits of this full-time grouping were essentially zero. In other words, it made no difference on children's ultimate achievement whether they were grouped full-time in high, average, or low level groups for their reading, math, or other instruction or if they were mixed heterogeneously in one classroom (currently referred to as *mixed-ability classrooms*). Kulik and Kulik (1982; 1984), however, reported a mean effect size of .15 for secondary students and .19 for elementary students when they combined the same studies used by Slavin with ability-comparison studies (studies in which children were placed in an ability group by their general intelligence level and differential effects of achievement were measured by comparisons with children at identical intelligence levels left in mixed-ability/traditional classrooms). The higher effect sizes of Kulik and Kulik might suggest that general levels of intelligence are an important consideration in the group-placement process if maximum achievement effects are to be realized for students at the highest intelligence levels. The Kuliks found that when intelligence levels were taken into account the ability comparison studies for gifted students (those in this research with IQs > 130) showed the greatest gains when these students were grouped full-time with other gifted students, over the performance of their intellectual peers who remained in traditional classrooms. Across the twenty-five studies that have looked at this question, the mean effect size for academic achievement for elementary gifted students was .49 (approximately one-half of a school year grade

gain) and .33 at the secondary level (one-third school year grade gain).

**THE MANAGEMENT VERSUS
CURRICULUM ISSUE: DOES ABILITY
GROUPING MAKE THE REAL
DIFFERENCE ALONE?**

It makes good sense to assert that it is probably not the practice of grouping by ability itself that makes a difference in students' achievement, but rather what goes on in the grouped situation. The rationale provided for decades has been that a smaller, more homogeneous group would naturally pave the way for more focused instruction, sensitive to the specific needs of the group. Yet in Slavin's studies, the effects of grouping where differences in instruction could be documented from group to group, no differences in achievement were noted (effect size = 0). The only group who seemed to make any gains in achievement in the grouped situation were the high achievers (in three-track studies) with an effect size of .12, a very small, positive gain, indeed (Slavin, 1987; Kulik & Kulik, 1990). Kulik (1992) summarized this situation:

Why were the effects of XYZ classes so small? The main problem with XYZ classes is probably their curricular uniformity. School personnel are usually careful in placing children into high, middle, and low classes, but they seldom adjust the curriculum to the ability level of the classes. For example, children in the high group in a Grade 5 program may be ready for work at the sixth grade level; children in the middle group are usually ready for work at the fifth grade level; and children in the low group may need remedial help to cover fifth grade material. But all groups work with the same materials and follow the same course of study in most XYZ classes. XYZ programs are thus programs of differential placement but not differential treatment. (p. xii)

On the other hand, studies that have documented differentiated curriculum and instruction in ability-grouped classes (for example, Provus, 1960), showed superior achievement effects over studies in which curriculum differentiation could not be documented. The high achievers, when curriculum was differentiated, gained by an effect size of .79, average achievers

by .22, and low achievers by .15 in the subject of mathematics. It is probably safe to say, therefore, that the practice of grouping by ability itself does not produce substantial gains in achievement, but the practice does facilitate the differentiation of instructional strategies and curriculum. Also at issue is the potentially moderating variable of teacher personality and effectiveness; unfortunately, the literature has not addressed this issue directly in ability grouping studies.

**PERMANENT VERSUS FLEXIBLE
QUESTION: IS THERE SUCH A THING
AS TOO MUCH OR TOO LITTLE, TOO
RIGID OR TOO FLEXIBLE GROUPING?**

Two issues are at question here. First, are there differing effects upon achievement and self-esteem or attitude toward academics and learning when children are grouped for the entire school day as compared with part-time grouping for only portions or specific subjects within the school day? Second, are there differing effects for students who are placed fairly permanently within an ability track as compared to students who are moved temporarily from group to group according to their current progress or lack of it? For answers to both issues, one must look at the effect sizes reported for tracked studies and full-time placement in special programs as compared with effect sizes in part-time grouping strategies such as regrouping for specific instruction or cross-grade grouping. For the general population and discounting all ability-comparison studies, effect sizes would appear to be larger for the part-time grouping practices. Substantial-effect sizes were reported, particularly for cross-grade grouping in reading at the elementary level and small gains were reported across K-12 in a variety of subject areas, whereas the general tracking studies were reported at zero (Slavin, 1987). For the special programs, especially for gifted students (Kulik, 1992) and for special education (Carlberg & Kavale, 1980), effect sizes are comparable for full- and part-time options. For the gifted, in particular, the overall effect size for full-time special-program placement across twenty-five studies was .33, down considerably from the .65 reported for pullout groups and .79 for specific instruction in mathematics. It is suspected, however,

in this body of research that if the part-time studies were equivalent in number, the mean effect sizes would decline to similar levels of achievement gain.

The body of studies on self-esteem ($n = 14$) has focused primarily on short-term effects when placed in full-time ability grouped settings, with mean effect sizes hovering close to zero, although the low-achiever groups tended to make small, positive short-term gains, unlike the average and high achievers under study. Studies of short-term effects on attitude toward school and learning have also only been focused on full-time ability-grouping studies, with effect sizes at .37 for secondary students and .27 across the K-12 body of studies (Kulik & Kulik, 1982; 1990; 1992). The affective issues of attitudes and esteem notwithstanding, would the achievement differences, as reported effect sizes suggest, indicate that there can be too much of a "good thing" with ability grouping if it is done too frequently? This is a difficult question to answer on the basis of current documentation. Common sense might suggest, however, that moderation is the key to any educational decision, that a variety of strategies, of which ability grouping is an important one, and perhaps mixed ability and homogeneously grouped cooperative learning are others, would provide the best probability of meeting the diverse needs, learning styles, and academic and experiential differences in every 1990s classroom and school. Rigorous evaluation of each experiment would clarify the effects of each practice.

CONCLUSION

The implications of these findings are fairly clear: Grouping by ability level cannot be discarded as a viable instructional strategy in the classroom teacher's repertoire, but care must be taken to determine when and for how long it will be used and with which groups of students.

(See also: CLASSIFICATION OF INTELLIGENCE.)

BIBLIOGRAPHY

- ABRAMI, P. C., COHEN, P. A., & D'APPOLLONIA, S. (1988). Implementation problems in meta-analysis. *Review of Educational Research*, 58, 151-180.
- CARLBERG, C., & KAVALE, K. (1980). The efficacy of special versus regular class placement for exceptional children: A meta-analysis. *Journal of Special Education*, 14, 295-309.
- COHEN, J. (1977). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- GLASS, G. V. (1976). Primary, secondary, and meta-analysis of research. *Educational Researcher*, 5(10), 3-8.
- GEORGE, P. S. (1988). *What's the truth about tracking and ability grouping really???* Gainesville, FL: Teacher Education Resources.
- HEDGES, L. V., & OLKIN, I. (1984). Analyses, reanalyses, and meta-analysis. *Contemporary Education Review*, 1, 157-165.
- HORNBECK, D. (CHAIRMAN). (1987). *Turning points: Preparing American youth for the 21st century*. Washington, DC: Carnegie Task Force on the Education of Young Adolescents.
- KULIK, C-L. C. (1985). *Effects of inter-class ability grouping on achievement and self-esteem*. Paper presented at the annual convention of the American Psychological Association, Los Angeles, CA.
- KULIK, C-L. C., & KULIK, J. (1984). *Effects of ability grouping on elementary school pupils: A meta-analysis*. Paper presented at the annual meeting of the American Psychological Association, Ontario, Canada.
- KULIK, J. A. (1992). *An analysis of the research on ability grouping: Historical and contemporary perspectives*. Research-Based Decision Making Series. Storrs, CT: National Research Center on the Gifted and Talented.
- KULIK, J. A., & KULIK, C-L. C. (1982). Effects of ability grouping on secondary school students: A meta-analysis of evaluation findings. *American Educational Research Journal*, 19, 415-428.
- KULIK, J. A., & KULIK, C-L. C. (1990). Ability grouping and gifted students. In N. Colangelo & G. A. Davis (Eds.). *Handbook of gifted education* (pp. 178-196). Boston: Allyn & Bacon.
- OAKES, J. (1985). *Keeping track: How schools structure inequality*. New Haven, CT: Yale University Press.
- PASSOW, A. H. (1962). The maze of the research on ability grouping. *Educational Forum*, 26, 281-288.
- PROVUS, M. M. (1960). Ability grouping in mathematics. *Elementary School Journal*, 60, 391-398.
- ROSENTHAL, R. (1984). *Meta-analytic procedures for social research*. Beverly Hills, CA: Sage.
- SLAVIN, R. E. (1986). Best-evidence synthesis: An alternative to meta-analytic and traditional reviews. *Educational Researcher*, 9(15), 5-11.

- SLAVIN, R. E. (1987a). Ability grouping: A best-evidence synthesis. *Review of Educational Research*, 57, 293–336.
- SLAVIN, R. E. (1987b). Ability grouping and its alternatives: Must we track? *American Educator*, 32–48.
- SLAVIN, R. E. (1990). Achievement effects of ability grouping in secondary schools: A best-evidence synthesis. *Review of Educational Research*, 60, 471–499.
- VAUGHN, V. L., FELDHUSEN, J. F., & ASHER, J. W. (1991). Meta-analyses and review of research on pull-out programs in gifted education. *Gifted Child Quarterly*, 35, 92–98.
- WITTY, P. A., & WILKINS, L. W. (1933). The status of acceleration or grade skipping as an administrative practice. *Educational Administration and Supervision*, 19, 321–346.
- WOLF, F. M. (1986). *Meta-analysis: Quantitative methods for research synthesis*. Quantitative applications in the social sciences series. Beverly Hills, CA: Sage.

KAREN B. ROGERS

ABSTRACTION The ability to engage in abstraction—that is, to separate an object from its immediate context—has long been considered one of the hallmarks of human intelligence. The American psychologist Lewis TERMAN, one of the original developers of intelligence tests, defined intelligence as “the ability to do abstract thinking.” Charles SPEARMAN, the British psychologist who pioneered the concept of general mental ability, believed that intelligence involved the “eduction of relations” and “the eduction of correlates.” Both processes depend on an initial abstraction of the features of objects to educe how those objects might be related. Given its centrality in definitions of intelligence, it is not surprising that the concept of abstraction has been employed in a wide variety of contexts, including neuropsychological assessment, studies of human development, and programs on increasing intelligence and thinking skills.

What exactly is abstraction? *Abstraction* refers to the process by which humans categorize things (objects, numbers, events, people, places) into higher-order classes. For example, if we are shown an apple, we might categorize it as “fruit”; if we are shown a guitar, we might categorize it as “a musical instrument.” Abstraction is often a prerequisite for seeing relations among disparate objects. When presented with the

numbers 2 4 6, for example, we abstract that these are even numbers that are related to each other in a series. This abstraction allows us to produce the next number in the series, which we know because the next number will also be an even number and it will be the next higher even number after 6.

Abstraction, or abstract thinking, may be contrasted with concrete thinking, or the inability to think about something in a context other than that in which it is presented. One who can successfully think abstractly has no difficulty in flexibly thinking about things in ways other than the ways in which those things are presented. For example, when shown a map of the world and asked, “How are these countries—the United States, Canada, and Mexico—alike?” a fairly concrete answer might be something like “They are all close together on the map.” An answer requiring somewhat more abstraction would be something like “They are all North American countries” or “They are all countries in the western hemisphere” or “north of the equator.” If the questioner asked for something else, a concrete thinker might have a difficult time thinking about something other than features of the countries tied closely to what is seen on the map, such as their color, shape, or geographic features (“They all have coastlines on both sides”). One more capable of abstraction would have an easier time thinking about features of the countries other than those associated with the map. For example, such a person might think about how all three countries include a significant proportion of people who speak both English and Spanish or how the population of all three countries includes a substantial number whose grandparents were born in another country or even how all three countries have three vowels in each word of their names or have three syllables.

TESTS OF ABSTRACTION

Abstraction is necessary for the solution of many different kinds of intelligence-test items, such as analogies (mother is to father as sister is to ___?), sets (Choose the odd set: fhj aci moq suw.), series (a c e g –), and matrix items. For various historical reasons (which might or might not be supported by solid empirical evidence), certain tests have developed a reputation as being particularly good tests of abstraction.

One test widely used to measure abstraction is the “similarities” test found on the various Wechsler intelligence scales. The similarities test consists of a series of word pairs such as *orange–banana*, *wagon–bicycle*, *wood–alcohol*, and *fly–tree*. An examiner asks an examinee how the two words in the pair are related. Examinees who point out the higher-level category in which the two terms belong, such as “Both are fruit” for *orange–banana* or “Both are means of transportation” for *wagon–bicycle*, are given full credit; those who point out other features, such as “You can eat them” or “They have wheels” are given only partial credit (Matarazzo, 1972). An interesting feature of this test is that the words used are all simple ones, known by most, if not all, examinees given the test. The difficulty is in the relationship between the two words rather than in the words themselves.

Another example comes from the Shipley Institute for Living Scale (SILS), a test developed in the late 1930s by Walter Shipley (1940) and still widely used in neuropsychological assessment. The scale consists of two tests, one measuring vocabulary and the other measuring abstraction. The abstraction test consists of twenty series items, similar to, but more varied than, the standard number-series items one finds on many tests of intelligence. An example of a rather easy item is *escape scape cape —*. An example of a more difficult item is *Scotland landscape scapegoat — ee*, and still more difficult, *two w four r three —*.

NEUROPSYCHOLOGICAL ASSESSMENT

Studies conducted on tests of abstraction have revealed two important features of such tests. First, abstraction tests are typically excellent measures of intelligence. Scores on both the similarities test and the abstraction scale from the SILS are among those most highly correlated with other measures of general intelligence. Second, the difference for individuals between their abstraction score and other scores from the battery, such as their vocabulary score, has been in some cases shown to be related to the onset of organic brain damage. In fact, the original motivation for the development of the SILS battery (Shipley, 1940) was to measure mental impairment. Early research had shown that individuals suffering from cortical injury, senility, schizophrenia, and other impairments showed

a marked loss in their ability to do classification and sorting tasks requiring abstraction, even when they showed no performance loss on other tests, such as vocabulary.

Partly for this reason, more recent neuropsychological investigations have routinely included tests of abstraction as potential correlates of a wide variety of conditions affecting cognitive performance. These have included studies on dementia and Alzheimer’s disease (Aniskiewicz, 1987), AIDS (Butters et al., 1990; Grant & Heaton, 1990), substance abuse (Meek, Clark, & Solana, 1989; Sweeney et al., 1989), learning disabilities (Scott, Greenfield, & Sterental, 1986), aging (Abraham & Reel, 1992), and Parkinson’s disease (Cummings, 1988). The point is not that there are firm conclusions regarding the loss of abstraction in relation to the onset of these various conditions but rather that neuropsychological investigators have deemed the possibility important enough to warrant the routine inclusion of abstraction in such studies. It might be pointed out, in this regard, that the status of abstraction as a distinct psychological construct appears to be more widely accepted within the field of neuropsychology than within the field of human intelligence in normal populations. Within the human-intelligence literature, one is more likely to encounter references to “reasoning ability,” for example, than to “abstraction ability.”

DEVELOPMENTAL PSYCHOLOGY

Theories of human cognitive development as it occurs from infancy to adulthood typically ascribe an important role to abstraction as a major feature of development differentiating mature from immature thinkers. Consider the theory associated with the Swiss psychologist, Jean PIAGET, who suggested that intellectual development occurs through four stages. An initial “sensorimotor” stage, occurring before the age of 2, is characterized by concreteness, in the sense that the infant experiments with concepts by taking actions with concrete objects. Next, during a “preoperational” stage (2–7 years), the child begins to use language to form higher-order categories of things, such as “shoes.” Following this, during what Piaget called a “concrete operational” phase (7–11 years), children can abstract themselves from their own ego-centric perspective and can adopt the perspective of

other people. In a final, “formal operational” stage, children can engage in true abstract thinking, such as what is required for solving “nonsense” syllogisms: “Suppose that all dogs were able to sing; could your dog sing?” Clearly, these stages differ from one another in the degree to which they allow for the individual to engage increasingly in abstraction, to remove himself or herself mentally from the context in which objects are presented and thereby to perform increasingly powerful mental manipulations of those objects.

TEACHING THINKING

If the ability to engage in abstraction is central to intelligent behavior, then it makes sense that programs designed to teach intelligence or to teach thinking skills might teach abstraction skills as a way to achieve their goals. Indeed, as has been pointed out in a recent review (Nickerson, Perkins, & Smith, 1985), there have been a number of promising teaching-thinking programs that attempt to do just that. In Edith Neimark’s *Adventures in Thinking* (1987), a number of exercises are given for what she calls “detaching,” which involves “overcoming impulse,” “depersonalizing” (e.g., taking the role of another, such as one with diametrically opposed views), and “broadening through abstraction” (e.g., thinking through the implications of a legal principle, such as prohibiting discrimination on the basis of sexual orientation).

A common difficulty in teaching abstraction is that principles taught in one context may not transfer to another context. Perkins (1987) suggested that it is important not simply to allow transfer of skills to occur spontaneously but to demand that the learner engage in what he calls “mindful abstraction,” a deliberate, conscious effort by the learner to seek generalizations of the principles learned and applications to new situations. This kind of deliberate abstraction of principles has been shown to be important in transferring mathematical problem-solving procedures from one content domain to another (Bassok, 1990).

CROSS-CULTURAL CONSIDERATIONS

A question is why the ability to engage in abstraction is so central to a conception of intelligence. Is abstraction a necessary component of intelligence? Or

is abstraction just something that the West as a culture happens to value? Some writers have suggested that the importance of abstraction is partly dictated by the culture. Abstraction depends, to some degree, on having language and concepts by which things can be classified. Literate societies, such as the West, make use of abstractions in categorizing objects and activities. As a result, one’s ability to succeed in such cultures is at least partly dependent on one’s ability to engage in abstraction. But this may not be a necessary feature of successful coping in a culture. Psychologists who have examined preliterate societies have noted that they do not necessarily put such a high premium on abstraction. For example, Michael Cole and S. Scribner found that preliterate adults in some cultures cannot solve abstract syllogism problems involving the acceptance of initial premises that might not be true. They argue that this is not because these adults lack intelligence but rather because their culture places no value on this kind of abstraction.

CONCLUSIONS

A key sign of normal intellectual development into mature thinking is the ability to abstract—to classify, categorize, and generalize in increasingly flexible ways, to imagine features of things presented other than those intended, and to detach principles from the context in which they are learned. In Western culture, these abilities are acknowledged as necessary, if not sufficient, to warrant the label “intelligent.” It may be possible to teach these abstraction skills, to a certain degree, although methods of doing so are thus far imperfect. There is some evidence that the loss of the capability to engage in abstraction is tied to organic brain damage. Finally, even though the ability to engage in abstraction is considered an important part of human intelligence, abstraction is not generally recognized by human-intelligence researchers as a distinct psychological construct; that is, there does not appear to be any evidence for an abstraction ability distinct from other intellectual abilities. Human intelligence researchers are more likely to refer to abstraction as an explanatory principle than as an ability per se.

(See also: REASONING, DEDUCTIVE; REASONING, INDUCTIVE.)

BIBLIOGRAPHY

- ABRAHAM, I. L., & REEL, S. J. (1992). Cognitive nursing interventions with long-term-care residents: Effects on neurocognitive dimensions. *Archives of Psychiatric Nursing*, 6(6), 356-365.
- ANISKIEWICZ, A. S. (1987). Neuropsychological assessment of cognitive impairment in the elderly. In J. A. Blumenthal & D. C. McKee (Eds.), *Applications in behavioral medicine and health psychology: A clinician's source book* (pp. 95-113). Sarasota, FL: Professional Resource Exchange.
- BASSOK, M. (1990). Transfer of domain-specific problem-solving procedures. *Journal of Experimental Psychology Learning, Memory, and Cognition*, 16(3), 522-533.
- BUTTERS, N., GRANT, I., HAXBY, J., JUDD, L. L. (1990). Assessment of AIDS-related cognitive changes: Recommendations of the NIMH Workshop on Neuropsychological Assessment Approaches. *Journal of Clinical and Experimental Neuropsychology*, 12(6), 963-978.
- COLE, M., & SCRIBNER, S. (1974). *Culture and thought: A psychological introduction*. New York: Wiley.
- CUMMINGS, J. L. (1988). Intellectual impairment in Parkinson's disease: Clinical, pathologic, and biochemical correlates. *Journal of Geriatric Psychiatry and Neurology*, 1(1), 24-36.
- GRANT, I., & HEATON, R. K. (1990). Human immunodeficiency virus-Type 1 (HIV-1) and the brain. Special Series: Acquired Immune Deficiency Syndrome. *Journal of Consulting and Clinical Psychology*, 58(1), 22-30.
- MATARAZZO, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore, MD: Williams & Wilkins.
- MEEK, P. S., CLARK, H. W., & SOLANA, V. L. (1989). Neurocognitive impairment: The unrecognized component of dual diagnosis in substance-abuse treatment. *Journal of Psychoactive Drugs*, 21(2), 153-160.
- NEIMARK, E. D. (1987). *Adventures in thinking*. San Diego, CA: Harcourt Brace Jovanovich.
- NICKERSON, R., PERKINS, D. N., & SMITH, E. (1985). *The teaching of thinking*. Hillsdale, NJ: Erlbaum.
- OLSON, D. R. (1976). Culture, technology, and intellect. In R. Resnick (Ed.), *The nature of intelligence*. Hillsdale, NJ: Erlbaum.
- PERKINS, D. N. (1987). Thinking frames: An integrated perspective on teaching cognitive skills. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and Practice* (pp. 41-61), New York: W. H. Freeman.
- SCOTT, M. S., GREENFIELD, D. B., & STERENTAL, E. (1986). Abstract categorization ability as a predictor of learning disability classification. *Intelligence*, 10(4), 377-387.
- SHIPLEY, W. C. (1940). A self-administering scale for measuring intellectual impairment and deterioration. *Journal of Psychology*, 9, 371-377.
- SPEARMAN, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.
- SWEENEY, J. A., MEISEL, L., WALSH, V. L., & CASTROVINCI, D. (1989). Assessment of cognitive functioning in poly-substance abusers. *Journal of Clinical Psychology*, 45(2), 346-351.
- TERMAN, L. M. (1916). *The measurement of intelligence*. Boston: Houghton-Mifflin.
- WECHSLER, D. (1939). *The measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins.

PATRICK C. KYLLONEN

ACHIEVEMENT TESTING Achievement tests are certainly the most common and are probably the most useful type of test in existence. Literally millions of achievement tests are given in the United States each year. Achievement tests are in use every time a classroom teacher gives a quiz, a college professor gives a final examination, a new driver demonstrates competence before receiving a license, a school district assesses its fourth-grade students' reading comprehension, a state requires high school students to demonstrate mastery of basic skills, or an employer asks a prospective secretary to take a typing test.

This article discusses achievement testing with a primary emphasis on the types of achievement testing done in schools in the United States. The entry defines achievement tests, discusses the differences between achievement tests and ability tests, describes the major uses of achievement tests, and gives a very brief sketch of the history of achievement testing in the United States. In addition, it outlines the general process used to construct achievement tests and stresses the importance of standardization. Finally, it describes the methods of attaching meaning to achievement test scores, norm-referencing, and criterion-referencing.

ASPECTS OF ACHIEVEMENT TESTS

Definition of Achievement Test. An achievement test is simply a more or less systematic sample of a person's behavior used to draw inferences about what the person currently knows or can do. The inferences can be in either absolute terms (can type forty words a minute) or comparative terms (is an average typist).

Differences Between Aptitude and Achievement Tests. A great deal of confusion exists about the distinction between achievement tests and so-called aptitude or ability tests. In one sense, all tests are achievement tests because all that any test can measure is what the test taker has achieved (and is willing to demonstrate) by the time the test is given. Tests can measure the test taker's behavior only at the time of testing. No way exists to measure innate ability, potential, or aptitude directly, uncontaminated by the influence of the test taker's environment and learning opportunities.

Not clearly distinct, as many people used to believe, ability and achievement tests overlap a great deal in both content and use. In content, the very same types of questions often appear in both kinds of tests. It is impossible, for example, to distinguish between an arithmetic question in a mathematics achievement test and an arithmetic question in a quantitative ability test, or between a reading comprehension question in a verbal ability test and a reading comprehension question in a reading achievement test. Achievement tests do, however, tend to relate to more recently taught material.

With respect to use, when a test examines ability, inferences are generally made about the test taker's future behavior. When tests examine achievement, inferences are generally made about the test taker's current levels of knowledge and skill. That distinction between the two types of tests is not strict. Some achievement tests have predictive purposes, as when a test of the knowledge of specific subject matter gained in college is used to predict success in graduate work in the same field.

The most important difference between achievement and ability tests is the assumption about the test taker's opportunities to learn whatever the test measures. Whenever tests examine ability, an underlying,

often unstated, assumption exists. Supposedly the opportunity to learn the material being tested is so widespread and easily available that all test takers have had a reasonably equal opportunity to learn it. Only if that assumption is true is it safe to interpret differences in scores as differences in ability. On the other hand, there is a clear, often explicit, assumption that the material covered in an achievement test is based on some course of instruction or period of training. People who have had no exposure to the training or instruction are not expected to do well on the test.

Purposes of Achievement Testing. The major uses of achievement tests are management of instruction, placement, selection, accountability, evaluation, and certification of competence. The different purposes do tend to overlap, and sometimes a single test serves several purposes at the same time.

Management of Instruction. One of the most common and important uses of achievement tests is to help teachers do their jobs well. Information about what students know and do not know throughout an instructional sequence is obviously crucial for the efficient management of instruction at all levels. Such information is required in any attempt by a teacher to tailor instruction to an individual or to a class. Achievement test results can identify both students in need of extra help and those for whom the material would be redundant. The distribution of scores for a class can help a teacher determine the most appropriate level of instruction for the group. In short, a teacher who is planning and monitoring a sequence of instructional activities can use achievement tests to help determine where it is best to start in the sequence with a class or with a particular student, how fast to cover the material so that students are challenged but not frustrated, and the appropriate place to stop because the objectives of instruction have been met.

Placement. To place students in courses covering material they have already mastered is clearly a waste of time and resources. Also wasteful (and somewhat cruel) is the placement of students in courses for which they lack the necessary background knowledge. Achievement tests can be useful in selecting which course in an ordered sequence would be most instructive for each student. This choice is accomplished by using a test to determine the student's particular level

of knowledge and skill and then comparing that level with the prerequisites and expected outcomes of each of the courses in the sequence. Placement tests are of most use in subject-matter areas that are hierarchical, in the sense that mastery of earlier objectives is necessary for the mastery of later objectives. Mathematics and foreign languages are clear examples of such subject-matter areas. Placement tests are also useful when different levels of the same course are available, as when a choice has to be made between an "honors" course and a standard course.

Selection. Achievement tests can also be very helpful in selection, because students who have done well at one level of schooling are the ones most likely to do well at succeeding levels. Selection differs from placement in that the selection of some applicants implies the rejection of others. Consequently, it is generally agreed that tests used for selection should meet higher standards of quality than tests used for placement.

An important reason for using standardized achievement tests for selection is that content coverage and level of rigor within courses of the same title may vary greatly across schools. Furthermore, grades cannot be compared fairly from school to school or even from teacher to teacher because of differences in grading standards. A grade of "C" in one school or from one teacher may represent the same level of knowledge and ability as a grade of "B" or even "A" in another school or from another teacher. Standardized test scores, however, can be compared fairly across teachers and across schools because all test takers are faced with the same tasks under the same conditions regardless of the location of the test.

Just as grades are limited in comparability, tests are limited in what they measure. Achievement tests cannot measure study habits, motivation, class participation, and completion of assignments, which tend to be reflected in grades. Therefore, the combination of test scores and grades offers more predictive power than either one used alone.

Accountability. Much public money goes to education, and legislators and their constituents have a right to know how well the schools are performing their function. Achievement tests provide a cost-effective, efficient, consistent, trustworthy means of demonstrating the accomplishments (or lack thereof) of

large groups of students within a school district or a state or across the country. In fact, international assessments can be done using achievement tests that have been specially constructed to maintain comparability of results in different cultures and when translated into different languages. Achievement tests also provide for accountability on the individual level. Combined with other factors such as classroom participation and completion of assignments, achievement tests are an important component of most school and college grading systems.

The large-scale use of standardized achievement tests for purposes of accountability, often driven by legislation, has become controversial. Critics contend that the tests have an inordinate influence on curricula and the allocation of instructional time, and that the widespread use of multiple-choice questions inhibits the teaching of higher-order thinking skills. When test results are used for the purposes of accountability, it is necessary to keep in mind factors that affect test results over which the school, the teacher, or the student has no control. Achievement test results must be interpreted with full knowledge of the context in which they were obtained, and care must be taken to avoid misuse and overinterpretation of the scores.

Evaluation. Achievement tests are invaluable in educational research concerning the effects of different curricula, texts, and teaching methods. The tests provide very useful information about how the variables being evaluated affected student learning. Achievement tests are, of course, limited in that they measure only what students have learned and not how students feel about their experiences. An educational treatment in which students learned the material well, but did not become motivated to study the subject again, would hardly be appropriate. Achievement tests are, therefore, necessary but not sufficient in the evaluation of educational treatments.

Certification of Competence. Achievement tests serve an important role in protecting the public when the tests determine whether people are qualified to perform certain activities. The most common example of this use of achievement testing is a driver's license examination. New drivers must demonstrate that they know traffic laws and that they are able to operate a vehicle safely. Achievement tests are useful

in various jurisdictions to determine whether licenses to practice should be awarded to teachers, lawyers, physicians, psychologists, building inspectors, realtors, insurance agents, stockbrokers, pilots, social workers, speech therapists, and many others.

A more controversial use of achievement tests has been to determine whether students have mastered basic skills well enough to receive a high school diploma. Many states instituted the tests because of complaints by employers and members of the public that too many high school graduates were illiterate, and that the diploma had lost all meaning as a guarantee of competence. The tests became controversial primarily because of different passing rates among students of various racial and ethnic groups. Some people see the tests as discriminatory, serving both to narrow the curriculum and to increase the number of dropouts. Others believe that the tests motivate students, give direction to teachers, and add meaning to the diploma.

A BRIEF HISTORY OF ACHIEVEMENT TESTING IN THE UNITED STATES

Although large-scale, sophisticated achievement testing originated in China over 3000 years ago, relatively unstructured oral examinations predominated in the United States until about the middle of the nineteenth century. Then rapid population growth, accelerated urbanization, and a corresponding increase in the numbers of students enrolled in public education forced schools to become more efficient to cope with the changes. Formal written testing proved increasingly useful for both instructional management and accountability purposes. World War I saw the development of mass testing and the large-scale use of multiple-choice questions by the armed forces to classify recruits, and the new technology was soon adopted for use in education.

By the middle of the twentieth century, the invention of the scoring machine had made objective achievement tests remarkably efficient and cost-effective, paving the way for the ubiquitous use within school systems of nationally standardized tests that were heavily dependent on multiple-choice questions. Federal and state mandated testing for evaluation and

accountability served to make standardized testing even more widespread. The increasing accessibility of computers for the analysis and reporting of scores allowed even greater gains in efficiency, utility, and cost-effectiveness and further augmented the growth of large-scale standardized achievement testing.

Toward the end of the twentieth century, advances in psychometric theory and in the power of personal computers led to the introduction of computer-based, adaptive testing, which allows the use of innovative types of nonmultiple-choice questions and offers even greater efficiency through individualized measurement.

Construction of Achievement Tests. The construction of most professional, large-scale achievement tests involves a similar series of developmental steps. The first step is to set the test specifications, the “blueprints” detailing the knowledge and skills to be tested. The judgments of representative groups of subject-matter experts are generally of great importance in determining the test specifications. For tests related to academic subjects, surveys of the contents of major textbooks and surveys of curricula help inform the experts’ judgments. For tests related to occupations, job analyses detailing the tasks required to perform the job are useful.

Once the specifications have been set and reviewed by samples of educators and job incumbents, test questions are written to meet the specifications. Typically, the questions are reviewed and edited a number of times to assure their quality. An important aspect of modern reviews is an explicit check for fairness to all test takers regardless of factors such as sex, race, ethnicity, and location. Generally, many more questions are written than are necessary, to allow for attrition during the development process.

Questions acceptable to the reviewers are administered to representative test takers in “pretests” to determine empirically the difficulty and measurement power (discrimination) of the questions. In modern test development, analyses of the differences in difficulty of the questions for various groups of test takers are carried out as an empirical check on the fairness of the questions. Tests are assembled from pools of questions with appropriate difficulty, discrimination, and fairness indices. Subject-matter, editorial, and fairness reviews are generally carried out on the draft tests. Rules for standardized administration and scor-

ing are generated, and the tests are then “normed” by administration to well-selected samples of test takers. Finally, after a process that can take two or more years, the tests are ready for use.

Standardization. An important feature of many achievement tests used with large populations is that they are “standardized.” In the context of testing, STANDARDIZATION means that all of the people taking the same test (with the exception of some people with disabilities) do so under the same conditions and that the scores are derived using the same rules for all test takers.

Standardization allows the results of achievement tests to be given meaning by comparing test takers’ scores to the scores of some relevant group(s). Without standardization, comparisons among test takers would be neither sensible nor fair. Reasonable comparisons would not be possible, for example, if some people were given two hours to take a test and others only one hour to take the same test, if some people were given two points for each correct answer and others only one point, or if some people could use calculators on a math test and others could not.

Therefore, standardization is a basic requirement for fairness and utility. Even for tests that report absolute rather than comparative scores, standardization is still necessary. Consider a road test for a driver’s license in which the test taker has to park a car. It would not be fair if one person had to parallel park in a tight space on a congested street and another person could park in an empty lot.

Types of Achievement Test Scores. Achievement test scores in isolation carry little or no meaning. The most widely used methods of attaching meaning to test scores depend on comparisons. In norm-referenced testing, a score is compared to the scores of some defined group of people. In criterion-referenced testing, the score is compared in some absolute passing score or standard. Test users find that both types of comparisons are necessary to obtain a complete description of a person’s achievement.

Norm-Referenced Scores. Many achievement tests become particularly useful because they are given under standardized conditions to carefully selected samples of the students in particular grades in a school district, a state, or across the country, and information about the resulting distributions of scores is collected

and analyzed. These scores can then be compared with the scores of other test takers. The sample groups are known as *norm groups*, and the scores are *norm-referenced scores*.

A very effective way to attach meaning to a test score is to compare it with the scores of some group(s) of people. For example, knowing that a child answered twenty-seven mathematics questions correctly indicates nothing about the child’s achievement in mathematics. Knowing, however, that the average child in the United States in the same grade answered seventeen of the questions correctly allows a clear inference that the child’s mathematics achievement is above average. Learning that 95 percent of the children in the same grade in the United States received lower scores clarifies the achievement even more. Additional comparisons may be made at other levels to gain even more information about the student’s relative standing and about the performance of various administrative levels of schooling in relation to others.

Although some norms are called *national*, no test publisher can possibly test every child in the country. Norms derive from samples of scores, and the sample obtained by a particular publisher may be more or less representative of the entire population. The sheer number of people in a norm group is of less importance than how representative they are. The norm group should replicate all the important characteristics of the target population in the same proportions, and the results of norm-referenced tests should be interpreted in the light of full information about the characteristics of the norm group(s) in use.

Criterion-Referenced Scores. Criterion-referenced scores have meaning in absolute, not comparative, terms. Rather than indicating, for example, that a student’s performance exceeds that of 36 percent of her peers, a criterion-referenced score may indicate whether the child can multiply fractions. The evaluations compare the child’s score with some passing score or standard. If a test includes ten questions on the multiplication of fractions, for example, a decision has to be made about how many of the questions must be answered correctly before the student can qualify as having mastered the skill of multiplying fractions.

A major problem with criterion-referenced scores is the setting of an appropriate standard. All methods of setting standards depend, at some point, on the sub-

jective judgments and values of the people who set the standard. Relatively high standards will reduce the chances of passing people who deserve to fail. Such high standards will, however, increase the chances of failing people who deserve to pass. On the other hand, relatively low standards will reduce the chances of failing people who deserve to pass. Yet such low standards will increase the chances of passing people who deserve to fail.

These "errors of classification" do not occur because some participant in the process made a mistake. Errors of classification occur because no test score can be perfectly accurate or precise. It is impossible to raise or lower a standard to reduce the likelihood of one of the types of errors of classification without simultaneously increasing the likelihood of the other type. (This is true regardless of the test or the method used to set the standard.) Because people have different opinions about which type of error is worse, no standard is likely to be equally acceptable to all concerned. The results of criterion-referenced tests should be interpreted in the light of full information about the characteristics of the standard(s) in use.

CONCLUSION

Achievement tests provide information about what people know and are able to do, generally as outcomes of specific, structured learning experiences. Achievement tests range from a surprise quiz on a chalkboard to nationally standardized examinations covering a number of subjects across all school grades. Large-scale achievement testing has become so pervasive and important in American education that some critics believe the tests to have undue influence and negative consequences. The critics propose the use of performance measures, locally developed measures, and portfolios of students' work to replace nationally standardized, multiple-choice tests. Almost all critics have proposed the replacement of one type of achievement test by other types rather than their abolition. Because achievement tests provide obviously useful information, running an educational system without some type of achievement testing is very difficult to imagine.

(See also: GROUP TESTS; SOCIALIZATION OF INTELLIGENCE.)

BIBLIOGRAPHY

- ANASTASI, A. (1988). *Psychological testing*. New York: Macmillan. A very useful, clearly written introductory text with broad coverage and lucid explanations of testing terminology. Contains an outstanding discussion of the distinction between ability and achievement tests.
- CONOLEY, J., & KRAMER, J. (EDS.). (1989). *The tenth mental measurements yearbook*. Lincoln: University of Nebraska Press. Part of a series of yearbooks first published in 1938. Begun under the editorship of Oscar Buros, the *Mental measurements yearbook* has long been the best source for evaluative information about particular tests. Requires some knowledge of testing terminology to understand the test reviews.
- LINN, R. (ED.). (1988). *Educational measurement* (3rd ed.). New York: Macmillan. An advanced text. Of particular interest with respect to achievement testing are chapters on validity, test bias, the specification and development of tests, administrative uses of school testing programs, certification of student competence, and educational admissions and placement.
- NITKO, A. (1983). *Educational tests and measurement: An introduction*. New York: Harcourt Brace Jovanovich. Easily understood general introduction to testing, with good sections on how to write the various types of test questions found in achievement tests.
- U.S. CONGRESS, OFFICE OF TECHNOLOGY ASSESSMENT. (1992). *Testing in American schools: Asking the right questions*. Washington, DC: U.S. Government Printing Office. Examines the federal role in testing and pays a great deal of attention to alternative means of assessment such as performance measures and portfolios. Has interesting information on testing in other countries and on the history of testing in the United States.
- U.S. GENERAL ACCOUNTING OFFICE. (1993). *Student testing: Current extent and expenditures, with cost estimates for a national examination*. Washington, DC: U.S. Government Printing Office. Detailed information on the extent, benefits, and costs of testing in United States schools.

MICHAEL ZIEKY

ADAPTIVE BEHAVIOR In its relationship to intelligence, the concept of adaptive behavior, while not new to psychology, achieved major influence as

a psychological construct and as an assessment dimension through its adoption by the then American Association on Mental Deficiency (now American Association on Mental Retardation). As a new dimension "in the measurement of mental retardation," *adaptive behavior* was defined three decades ago as "the effectiveness with which the individual copes with the natural and social demands of the environment" (Heber, 1961). Since that time the criteria that define adaptive behavior have been organized, systematized, and developed into a variety of scales and planning concepts, and used as a specific part of the definition of mental retardation and developmental disability. However, the process of adaptation is basic to the growth and development of all organisms.

The human must be considered essentially as a special adaptive organism. The human organism is differently adaptive than other living organisms in that it must consciously plan for its own survival. The human faced with an environment that contains other living organisms, all struggling for survival, must find its own way through the utilization of the surrounding milieu, and through adjustment of and accommodation to those environmental forces. The process of adaptation, therefore, is a very delicate balance that is achieved through active interaction with the immediate environment (Leland, 1977). The human must discern and select from that environment the cues and behavioral guides critical to the successful comprehension of the demands of that environment, and having realized these demands it must then, through its own processes, adjust its behavior and modify its approaches to develop individual strategies to deal with those demands. In short the human has evolved coping behavior. This human evolution of coping behavior becomes part of what might be described as an organic concept of intelligence (Leland, 1977).

Historically, the underlying concept has been around since philosophers first started recording their speculations about the nature of people. Seeing that individuals behave differently from their counterparts in the animal kingdom led to very early labeling of differences among those individuals to define these different behaviors as part of a total system. Both Plato and Aristotle had discussions on these types of human differences (e.g., the different characteristics needed to be a soldier, worker, or statesperson) and both had

very strong opinions as to the social value of individuals who behaved differently (Beare, 1906). This emphasis on adaptive differences became even more involved over the next one thousand years, as the Renaissance and the Reformation grew together in an uneasy symbiosis creating varied complexities of thought throughout western Europe. Supernaturalism and other-worldliness inherited from the Middle Ages combined with the Renaissance humanist concern for people in this world to produce subtle shifts of emphasis and various blendings of heterogeneous elements that finally evolved into a scientific methodology (Kagin, 1968).

In 1672, Willis attempted to develop one of the first systematic sets of definitions, in areas that would today be called psychology. He used specific behaviors to define the differences between typical clinical labels (Cranefield, 1961). Various texts describing efforts to deal with multiple problems of behavior, starting with the 1806 *Wild Boy of Aveyron* (Itard, 1962), were based on elements that are typically used today to describe adaptive behavior. However, the first recorded measurement of adaptive behavior was probably Felix Voisin (1843), who attempted to devise a scale that included measurement of differences among individuals in eating habits, friendliness, courage, disruptive tendencies, egoism, and mental dexterity.

In more modern usage, references to adaptive behavior are very culture specific. Humans are psychobiological organisms and each of us develops from and behaves within some sort of cultural social background. We are each the result of reciprocal interaction. Reciprocal functioning, of which adaptive behavior is a segment, along with social awareness, is the product of an interaction between cultural mores and biological, social, and personally experienced factors. The differences in the levels of ability shown by different individuals at each of these levels is dependent on the manner in which the individual copes with this interrelationship of forces. The adaptive ability to cope is part of intelligent behavior. Today we would say that it is the force that drives or it is the motivation of intellectual behavior (Leland, 1990).

We recognize that intelligence is dependent on a wide variety of reciprocal interrelational elements: (1) sensorimotor development; (2) cognition, which itself is subdivided into processes of reception, per-

ception, and apperception; (3) rate of learning; (4) adaptive behavior; and (5) social awareness (Leland, 1991). We have also come to recognize that learning in humans begins during gestational development, and that the baby, before birth, has already begun to acquire elements relating to the developing vestibular system; otic and optic systems; and the initial sensory phenomena surrounding them (Anokhin, 1974). Work that is being done on the orienting reflex, and measurements of prenatal heartbeats indicate that when certain stimuli are introduced to the fetus it responds in an orienting manner and it habituates (Berntson et al., 1985; Tuber et al., 1980). The fetus goes through a lot of sensory habituation during that period. Developmental errors can occur because of environmental pressures in the womb (e.g., maternal smoking, drugs, caffeine, alcohol), which may lead to imperfect development of the brain itself. Such developmental errors relate to failure of the brain to function appropriately after birth. This whole developmental package, starting prenatally, is part of the intellectual development in the individual. This is why it is agreed that the timing of the damage to the brain has more meaning than the actual nature of the damage (Luria, 1980). Thus the individual is not born a perfect blank at birth but rather is already a person with a working sensory system.

Going back a million years or so, we find that there has been considerable continuity in human development. Before we had evidence of human or hominoid presence, animal development gave us many of the same types of continuities: two eyes looking forward, two ears, one on each side, a mouth in front, a head on top, with a ratio of size of the cranial vault to the size of the rest of the body. Those we believe were the earliest humans evolved a short hip bone that was a very necessary adaptive mutation. Today we define the process of evolution as a process of inheritance, plus adaptation, plus social necessity (Sagan, 1977; Gould, 1983). There seem to be various types of behavior that are learned and some (e.g., the propensity of a newborn duck to bond itself to the first moving object it sees immediately after birth) that may not involve direct learning (Lorenz, 1965). Some of these learned behaviors or adaptive responses may result in maladaptive behavior (e.g., a duck attaching itself to a hu-

man as if it were the mother), which may be described clinically as pathological (Sidman, 1960).

The most important general discussion of the study of adaptation as a systematic and experimental approach to behavior in animals is contained in the work of H. Helson (1964). The leading element in his animal studies involved the manner in which the animal uses the surrounding environment to meet basic needs and to survive. Some animals use elements within the environment to enhance the quality of their life (e.g., use of support systems, or making natural objects work as tools). There also are very primitive examples, such as the amoeba or the cockroach, which seem to adapt by absorbing as much of the interfering environment as they can absorb into their systems. Larger animals scrape out space to work, build nests or lairs, use sticks to dig for clams, and in effect modify the environment to improve their chances of survival. A phenomenon described as "unlearned learning" has also evolved (e.g., the manner in which the body "learns" to regulate a heart beat, an eye blink, or the conversion of blood oxygen, etc.). One finds that there is a consistent pattern of development of regular and systematic functions in all internal organs under the heading of reflex responses, of which the adaptive reflex is major. For example, the eyelid must remain open to permit vision, but this puts it in danger of drying out, becoming painful and losing vision. The lid must close, but again, vision is lost. Intermediate, adaptive reflex responses known as the eye blink are established, which allow the eye to continuously open and close in a systematic manner as external pressures demand. From this example and other similar examples, it becomes clear that adaptive behavior is the result of the interaction of complex biological needs and complicated learning processes. Such behavior leads to further learning, through social experiences, mediated for the promotion of higher human mental function (Luria, 1980).

In all epochs of human existence there has been an ever greater necessity for coping with the environment and for creating modifications to provide both protection against the elements and protection from other organisms. To maintain and enhance the quality of life, there was a further need for the human to manipulate and modify the natural and human environment

(Coelho, Hamburg, & Adams, 1974). The growth that resulted has been a vast reservoir of learning and lore, based on both oral and written language, which has produced major modifications over time.

This evolution over time in human adaptive behavior has occurred directly through interaction with specific aspects of various environments. One result has been the creation of unique human groupings into cultural and subcultural organizations. Over time, there have evolved specific ways of doing things within a specific culture, and individuals performing differently may be derided as being wrong. Such questions as the proper ways to use utensils to feed oneself became representative of cultural differences that an outsider may not know. The usual or traditional response patterns within a given culture or subgroup give rise to two types of social demands; these may be described as necessary and desirable. The necessary demand is that persons feed themselves; the desirable demand is that they do it in a certain way, at a certain time, often in a certain place. These cultural (or national) differences become evident as more and more individuals come to live in closer proximity and respond differently to what appears to be the same stimulus.

There are, as a result, a wide range of culturally desirable behaviors that may not be really necessary for survival (e.g., certain modes of dress). These behaviors set up a high level of personal interaction and socialization and thus become extremely important in reciprocal communications. Many of these desirable behaviors may be more important, in the long run, than the more basic or necessary behaviors (Leland, 1982). Some parts of this new set of coping demands are now being researched and described as social competencies (American Psychological Association, 1992). They emphasize some of the steps in the utilization of personal social-historical experiences for social decision-making functions, with the uniquely human form of adaptive behavior emerging as one of the major mediating factors.

The study of adaptive behavior becomes the basis for increasing our understanding of how persons can bring themselves into active participation within their community in a manner that is judged to be of service. The community then rewards that service psychologically by allowing those persons to remain. Thus, J. V.

Wersch (1985) spoke of "the degree that a child grows into the surrounding social environment in the process of adaptation (as a measure) of how he or she develops the basic opportunities in individual sets of behavior." Lev VYGOTSKY (1986) did not attempt to measure the differences among individuals in their degree of adaptation. He outlined an adaptive hierarchy of cognitive growth: As cognition and intelligence develop, the intrapsychical growth modes emerge with greater individuality. Because people must live in union with each other, the interpsychical modes (speech, emotions, shared learning differences, etc.) become the major routes through which individuals create the best atmosphere for the utilization of adaptive techniques to cope with the expanding and differentiating roles of populations. These differing populations include differences in vast areas of life-style within and between units of the human family. They also include understanding of interrelationships between humans and nonhumans on all types of behavior, and the concept of necessary adaptive behavior takes in its full meaning as a mediating force of intelligence. This process of understanding is apart from the process of learning adaptive skills. Although they meet the requirements of desirable behaviors, adaptive skills also tend to become static and fixed in certain behavior patterns, and persons with similar coping skills can be described as *socially invisible*. They represent a group often described as *normal*. Both those persons with above average coping skills and those with below average coping skills are, by definition, abnormal (outside the norm) and socially visible. Members of the highly functioning group can use their visibility to aid their survival and are often honored and protected by society (e.g., artists, specialized professionals, university professors). Members of the lower functioning group are impeded by their social visibility and are often rejected by society (e.g., exploited in the institutions that house them or demeaned by classmates). In this context adaptation evolves as the reversible aspects of intelligence.

Adaptation, therefore, may be expressed in the form of successful (invisible) coping and can embrace social participation and social acceptances, or it may be expressed as unsuccessful (visible) coping, which sets the individual apart from society and forces society to decide whether to accept or not to accept the

person. Either process may be modified or reversed through social intervention (e.g., teaching, therapy, new experiences).

How do we measure these processes? Because we are dealing with adaptive coping, we have to start by defining the measurement in terms of subject behavior. DeJung (1963) points out that such an approach does not necessarily exclude evaluation of something someone is or has, such as being intelligent or experiencing emotions. Adaptive behavior is also concerned with those characteristics, but for the direct measurement of adaptive behavior we have to deal with the differences among individuals in real-life behavior that can be observed by others. Individuals behave in a variety of ways, and when a person's level of adaptive behavior is being assessed (by a psychologist, for example), they may not always behave alone as they do when the examiner is present (people do not behave consistently at all times). One has to consider what behavior is to be evaluated and under what circumstances one might expect that behavior to occur. Inasmuch as the typical examiner is not in a position to follow the human being examined for twenty-four to forty-eight hours at a time, to become aware of the various nuances of that person's behavior, staged or standardized observation of behavior is not a practical mode of testing. The examiner is dependent rather on other individuals (the mother of a child, the spouse of a patient who is recovering from a stroke) who have daily knowledge of the person being assessed and who can report to the examiner what differences have been observed over time, about particular coping mechanisms or adaptive abilities being evaluated. The instrument used to quantify that individual's level of adaptive behavior must be one whose reliability and validity have previously been established. In the development of such an adaptive behavior scale or instrument, validity must be established by observation of specific types of behavior that two or more observers agree upon.

For this reason, effective, valid measurement of adaptive behavior has been reduced to an observation of very specific behaviors that, of necessity, stand for a wide range of learning and coping abilities within broadly defined areas of social and personal interactions. Many of these observations have been reduced to either/or measurements with only minimal effort to measure variations. Thus, one asks only if an individual

child or recovering adult can walk, without concern about other features of such mobility, such as the nature of the gait. Higher levels of mobility beyond walking also receive points on such a scale (e.g., running and jumping). Other demands of such adaptive behavior measuring scales are less specific, having to do with cooperation, personal responsibility, or social behaviors, which imply levels of development. In measuring the current level of adaptive behavior, one is measuring a modifiable dimension of intelligence. Consequently, there are certain types of information that must be obtained by the examiner.

First, what are the cultural criteria for necessary or desirable behaviors among the members of the social group being evaluated? How does one person's behavior differ from another person's and what is the range of social tolerance for that difference? The interpretation of the results of the measurement may vary from group to group (e.g., a computer operator does not have to be able to hammer nails, but a carpenter does). Tests of adaptive behavior that try to imply too much from irrelevant tasks have little use in adaptive behavior measurement, regardless of how statistically reliable these tests are.

Second, because adaptive behavior is more than just a collection of skills (these scales should not be confused with ACHIEVEMENT TESTING), the criteria for behaviors should be part of an accepted hierarchy of intellectual growth and development. Thus, although the absence of certain behaviors considered adaptive does not necessarily imply lack of capacity, it does indicate that the individual is not responding in the expected manner (i.e., as do his or her age peers) to those stimuli in the environment. This leads to the important concept of hills and valleys. The hills represent the person's adaptive strengths and levels of learning from previous experience (and, by implication, areas of behavior where one might expect future success and future learning). The valleys represent current weaknesses and those areas that could well have produced adaptive failure in the past and/or where there were physical or psychological impediments to learning and successful coping.

The concept just described is very similar to the "zone of proximal development" introduced by Vygotsky (1986), which views adaptation as unifying, covering a number of unevenly developed areas of

intellectual growth. In counseling youngsters who appear to be slow in their adaptive skills, the function of measurement is to determine those hills that, with the aid of a teacher or sophisticated counselor, might continue to improve. This approach to learning assumes that the social demands are present, and if individuals fail to cope successfully (i.e., are exhibiting valleys in their adaptive repertoire), it is because they had no reference base to successful coping in that area. In other words, there was no adaptive readiness, which an effective adaptive scale profile should help identify.

Third, adaptive behavior measurement cannot exist in a social vacuum. Such responses change from setting to setting. With children there are often major differences in adaptive skill between home and school (Shaw, Hammer & Leland, 1991). With adults there are similar differences linked to changes in personal circumstances (Gunsett, 1985). Thus adaptive behavior is not static or fixed. It is rather a flexible, remediable index that reflects changing moods, behaviors, and self-direction.

The general purpose of adaptive behavior assessment is to establish a valid measure of the level of the individual's skill in current, as well as anticipated, behaviors in order to provide guidance for professionals working with persons with various ranges of mental and physical disabilities. Assessment is not aimed at traditional diagnostic labels comparable to one's IQ; it is rather a way of determining how a person may be expected to cope in average or usual daily activities versus unusual or unexpected social situations.

It is important to establish the types of behaviors that persons with disabilities are able to execute, in comparison with their peers of different levels of ability. Also, it is important, assuming that most other elements remain constant, to know in what areas of potential and appropriate remediation it will be possible for an individual to do more. This gives new meaning to prognosis, and to the relationship between prognosis and diagnosis.

(See also: CULTURE AND INTELLIGENCE.)

BIBLIOGRAPHY

AMERICAN PSYCHOLOGICAL ASSOCIATION. (1993). American Psychological Association, 100 Years: 1892-1992, Cen-

- tenial Convention Program—APA Science Weekend, *Increasing Competence and Adaptive Behaviors*, Washington, DC: Author.
- ANOKHIN, P. K. (1974). *Biology and neurophysiology of the conditioned reflex and its role in adaptive behavior*. Oxford: Pergamon Press.
- BEARE, J. I. (1906). *Greek theories of elementary cognition from Alcmaeon to Aristotle*. Oxford: Clarendon Press.
- BERNTSON, G. G., RONCA, A. E., TUBER, D. S., BOYSEN, S., & LELAND, H. (1985). Cardiac reactivity and adaptive behaviors. *American Journal of Mental Deficiency, 89*(4), 415-419.
- COELHO, G. V., HAMBURG, D. A., & ADAMS, J. E. (1974). *Coping and adaptation*. New York: Basic Books.
- CRANFIELD, P. (1961). A seventeenth century view of mental deficiency and schizophrenia: Thomas Wills on "Stupidity or foolishness" (1672). *Bulletin of the History of Medicine, 35*(4).
- DEJUNG, J. E. (1963). *An approach to the problem of measurement of adaptive behavior*. Paper presented at the South Central Regional Conference, American Association on Mental Deficiency, Enid, OK.
- GOULD, S. J. (1983). *Hens' teeth and horses' toes*. New York: W. W. Norton.
- GUNSETT, R. (1985). *Neuropsychological functioning and its relationship to treatment outcome in injured industrial workers with chronic pain*. Unpublished doctoral dissertation, Ohio State University, Columbus.
- HEBER, R. (ED.). (1961). A manual on terminology and classification in mental retardation, 2nd ed. (Monograph Supplement). *American Journal of Mental Deficiency, 65*.
- HELSON, H. (1964). *Adaptation-level theory*. New York: Harper & Row.
- ITARD, J.-M. G. (1962). *The Wild Boy of Aveyron (1806)*. New York: Appleton.
- KAGIN, E. F. (1968). Adaptive behavior and mental retardation during the Renaissance and Reformation. *Proceedings of the 76th Annual Convention*. Washington, DC: American Psychological Association.
- LELAND, H. (1977). Adaptation, coping behavior, and retarded performance. In P. Mittler (Ed.), *Research to practice in mental retardation: Education and training* (Vol. 2). London: University Park Press.
- LELAND, H. (1982). Assessment of adaptive behavior. In K. D. Paget & B. A. Bracken (Eds.), *The Psychoeducational assessment of preschool children*. New York: Grune and Stratton.

- LELAND, H. (1991). Intelligence and adaptive behavior revisited. *Psychology in Mental Retardation and Developmental Disabilities*, 17(1), 7–9.
- LORENZ, K. (1965). *Evolution and modification of behavior*. Chicago: University of Chicago Press.
- LURIA, A. R. (1980). *Higher cortical functions in man* (2nd ed.—rev.). New York: Basic Books.
- SHAW, J. G., HAMMER, D., & LELAND, H. (1991). Adaptive behavior of preschool children with developmental delays: Parent versus teacher ratings. *Mental Retardation*, 29(1), 49–53.
- SIDMAN, M. (1960). Normal sources of pathological behavior. *Science*, 132(3419), 61–68.
- TUBER, D. S., BERNTSON, G. G., BACHMAN, D. S., & ALLEN, J. N. (1980). Associative learning in premature hydronephalic and normal twins. *Science*, 210, 1035–1037.
- VYGOTSKY, L. (1986). *Thought and language* (1929). Cambridge, MA: MIT Press.
- VOISIN, F. (1843). *De l'Idiotie chez les enfants*. Paris: J. B. Baillière.
- WERSCH, J. V. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.

HENRY LELAND

ADAPTIVE BEHAVIOR, ASSESSMENT OF

Measures of adaptive behavior are designed to assess what an individual actually does in adapting to the requirements and challenges of living. They attempt to specify how well the person carries out the tasks of daily life and meets the social expectations of his or her environment. They also attempt to convert theories about human adaptive behavior to practical, psychometrically sound measures of individual differences. These measures are currently less advanced than analogous measures in the study of human intelligence. Less scientific attention has been devoted to adaptive behavior than to intelligence; not surprisingly, efforts to develop reliable and valid measures of adaptive behavior have lagged behind as well. The quality of available adaptive-behavior measures, however, has improved markedly in recent years.

HISTORICAL CONTEXT

During the 1970s and 1980s, scientific and professional interest in adaptive behavior burgeoned in response to a variety of social, legal, and political influences affecting the delivery of educational and social services to handicapped individuals. Especially influential were the changes in federal law (P.L. 94–142, 1975), changes in the legally accepted definition of mental retardation (e.g., Heber, 1961), and the findings of a number of highly publicized class-action lawsuits (e.g., *Larry P. v. Riles*, 1972, 1974, 1979; *PACE v. Joseph P. Hannon*, 1980).

Since 1975, federal law has required that adaptive behavior be considered *in conjunction with* intellectual functioning in the assessment of mental retardation (P.L. 92–142). This practice is congruent with the most recent American Association of Mental Retardation (AAMR) definition of mental retardation as “significantly subaverage intellectual functioning, existing *concurrently with* related limitations in two or more of the following applicable adaptive skill areas: communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academics, leisure, and work” (Luckasson et al., 1992, italics added).

HOW ADAPTIVE BEHAVIOR IS MEASURED

Unlike intelligence, adaptive behavior is almost never measured by “testing” an individual directly. Instead, a third party, someone who knows the individual well and is familiar with the person’s daily habits, is asked to rate the individual’s typical behavior in a variety of contexts. In a structured interview, the informant (usually a parent, teacher, or other caregiver) is asked such questions as “Does the individual use a knife and fork when eating?” The responses are then recorded in such categories as “usually,” “sometimes,” “never,” or “no opportunity to observe,” and are scaled a priori with assignment of numbers such as 2, 1, 0, or a missing-data code. The numbers are then added over several items and converted into standardized scores based on comparisons with one (or more) representative norm group(s). In this respect, adap-

tive-behavior scores are similar to intelligence or "IQ" scores.

RELATIONSHIP BETWEEN ADAPTIVE BEHAVIOR AND INTELLIGENCE

The relationship between adaptive behavior and intelligence has been studied fairly extensively, especially the question of whether adaptive behavior is sufficiently different from intelligence to be considered an entirely separate psychological construct. Generally speaking, a considerable overlap exists between people's intelligence scores and their adaptive-behavior scores. The amount of overlap between the two is directly influenced, however, by the way the construct of adaptive behavior is *operationalized*. In other words, the degree of overlap depends largely upon the particular "test" of adaptive behavior.

Research studies employing the Adaptive Behavior Inventory for Children (ABIC), a scale that intentionally excludes school-related behaviors (Mercer, 1979), consistently report minimal overlap; but studies employing more developmentally based or cognitively based scales typically report extensive overlap with intelligence. Most studies, however, indicate a moderate degree of overlap ($r = .40$ to $r = .60$), a pattern that is consistent with the hypothesis of two *separate but related constructs* (Platt et al., 1991).

Other findings relating adaptive behavior to intelligence indicate that (1) correlations between intelligence and adaptive behavior tend to be higher for more heterogeneous, or varied, subject pools and for subjects with lower mental ages; (2) adaptive-behavior scales that emphasize underlying cognitive or school-related skills yield higher correlations with intelligence than do scales that focus exclusively upon out-of-school adaptive skills; (3) adaptive-behavior ratings obtained through direct observation or through teacher ratings tend to correlate more highly with intelligence than do ratings obtained via parent ratings; and (4) certain specific domains within adaptive behavior measures are related differentially to intelligence scores; the communication-cognitive-skills domain tends to be most strongly related, and social adjustment is least strongly related (Bruininks & McGrew, 1987; Harrison, 1987, 1989; Lambert, 1981).

WIDELY USED SCALES OF ADAPTIVE BEHAVIOR

Several of the best-known and most widely used scales of adaptive behavior are described below.

AAMD Adaptive Behavior Scale for Children and Adults (ABS). This measure is intended to assess "the way an individual maintains his or her personal independence in daily living or how he or she meets the social expectations of his or her environment" (Nihira et al., 1975, p. 5). The scale is divided into two parts.

Part I includes ten adaptive categories:

1. independent functioning (eating, toilet use, cleanliness, appearance, dressing)
2. physical development (sensory and motor skills)
3. economic activity (money handling and shopping)
4. language development (receptive and expressive language)
5. numbers and time (use of number and time concepts)
6. domestic activity (e.g., cleaning, kitchen duties, laundry)
7. vocational activity (job performance and work habits)
8. self-direction (initiative, perseverance, and use of leisure)
9. responsibility (care with personal belongings and general reliability)
10. socialization (appropriate and inappropriate social behaviors).

Part II includes fourteen maladaptive categories:

1. violent and destructive behavior (e.g., temper tantrums, property damage)
2. antisocial behavior (e.g., swearing, teasing, bossiness)
3. rebellious behavior (e.g., disobedience, noncompliance, running away)
4. untrustworthy behavior (lying and stealing)
5. withdrawal (shyness, inactivity, and reticence)
6. stereotyped behavior and odd mannerisms (tics and unusual personal habits)
7. inappropriate interpersonal manners
8. unacceptable vocal habits

9. unacceptable or eccentric habits (e.g., removing clothing)
10. self-abusive behavior
11. hyperactive tendencies
12. sexually aberrant behavior (e.g., public masturbation, rape)
13. psychological disturbance (e.g., hypochondriacal tendencies, low frustration tolerance, excessive need for attention)
14. use of medication.

The norms for this instrument, based on ratings from approximately 4,000 institutionalized individuals, are seriously flawed. They are now outdated. They include no normally functioning individuals. They also appear to be identical to the norms used in the 1969 version of the ABS, even though the 1974 edition includes several revised items. Furthermore, they suggest that an individual tested at successive ages is losing ground, even when the level of adaptive behavior has remained exactly the same.

The only type of reliability addressed in the ABS manual is interrater reliability—the extent of agreement between different informants. These reliabilities range from a high of $r = .93$ in physical development to a low of $r = .37$ (inadequate for most purposes) in unacceptable vocal habits. The extents to which the different items reliably agree (i.e., internal consistency reliability) and measures remain stable over time (i.e., internal consistency reliability) are adequate for Part I but inadequate for Part II.

Evidence that the ABS is a valid measure of adaptive behavior (i.e., indications that it measures what it is intended to measure) is weak, especially for the domain scores. The ABS domains are never clearly defined, and the criteria for selecting items to represent a particular domain are not specified. Several domains in Part II contain only one or two items each. In addition, items scored in the negative direction (e.g., homosexuality and masturbation) would not necessarily be viewed as maladaptive by general consensus. In the category “use of medication,” individuals who take medications for seizures or hyperactivity are arbitrarily scored as less adaptive than those who do not.

AAMD Adaptive Behavior Scale, School Edition. The AAMD Adaptive Behavior Scale School Edition (ABS-SE; Lambert & Windmiller, 1981) is a

modified version of the ABS for use with children in the public schools. The scale’s stated purpose is to provide information about a student’s “personal independence and social skills and to reveal areas of functioning where special program planning” would be beneficial (Lambert et al., 1981, p. 3).

Three ABS domains deemed nonapplicable to school settings (domestic activity, self-abusive behavior, and sexually aberrant behavior) were not retained in the ABS-SE. The content of the two scales is essentially the same otherwise; the difference consists of minor changes in the wording of some items.

The informant for the ABS-SE is usually a teacher or a parent; ideally, both may be interviewed and their ratings compared. Percentile ranks are provided for each of the various domain scores and for a single summary score. Scores may also be converted to five broad category (factor) scores: personal self-sufficiency, community self-sufficiency, personal-social responsibility, social adjustment, and personal adjustment.

The norm group for the ABS-SE included 6,500 children from California and Florida only, some from regular education classes and some from special education classes for mildly and moderately mentally retarded students. No information is available regarding the balance of this sample. Separate norms are provided for comparison with students at various age levels, and at various levels of educational placement (both regular education classes and special education classes for mildly and moderately retarded students). Unfortunately, some of the samples on which these norms are based were clearly too small to be representative.

No information about the reliability of domain scores or composite scores is available; analysis of the broader category (factor) scores indicates, however, that the items within each category are consistently measuring the same type of behavior. According to the manual, the relationship between Part I scores and IQ is generally moderate. Part II scores showed no significant relationship to IQ, however.

Adaptive Behavior Inventory for Children (ABIC). The Adaptive Behavior Inventory For Children (ABIC; Mercer & Lewis, 1978) was designed for children aged 5 to 11 years, 11 months. Published as one component of the System of Multicultural Pluralistic Assessment (SOMPA; Mercer, 1979), the ABIC

may be administered in English or in Spanish, preferably through an interview with the child's principal caregiver. Six different domains are included: family, peers, community, school (nonacademic behavior only), earner-consumer, and self-maintenance. Unlike other adaptive behavior scales, the ABIC specifically excludes items that measure the more cognitive or academic aspects of school behavior.

The norm group included 2,085 children from California only, with equal numbers of boys and girls at each age level and approximately equal numbers of children from each of three ethnic groups. Reliability coefficients for both the total sample and for the three separate ethnic groupings are reported as high.

According to the SOMPA manual, the validity of the ABIC can be judged by its "ability to reflect accurately the extent to which the child is meeting the expectations of the members of the social systems covered in the scales" (Mercer, 1979, p. 109). This judgment is said to be accomplished by comparing "direct" measures of a child's adaptive behavior with "indirect" measures of the same behavior. ("Direct" measures are defined as evaluations that come from "members of the system . . . such as a peer evaluating a child's performance in the peer group"; "indirect" measures are defined as evaluations from individuals outside of the system [Mercer, 1979, p. 107]). Unfortunately, no comparisons between such direct and indirect measures are available either in the manual or in subsequent research.

The relationship between ABIC scores and intelligence scores of children in the norm group was extremely weak: $r = .16$ for Wechsler Verbal IQ, $r = .14$ for Performance IQ, and $r = .17$ for Full Scale IQ. These results, which have been corroborated by subsequent research, appear to reflect the intentional exclusion of cognitive/academic content in the design of this scale.

Scales of Independent Behavior (SIB). The Scales of Independent Behavior (SIB; Bruininks et al., 1984) are designed to evaluate adaptive behavior across the lifespan, from infancy through adulthood. Four adaptive categories are included: motor skills, social interaction and communication skills, personal living skills (e.g., eating, toilet use, dressing, domestic skills), and community living skills (e.g., work skills, punctuality, handling of money). Three maladaptive

categories are also included: internalized maladaptive behavior (hurtful to self, unusual or repetitive habits, and withdrawal or inattentive behavior), asocial maladaptive behavior (socially offensive behavior and uncooperative behavior), and externalized maladaptive behavior (destructive to property and disruptive behavior). Usually the informant is a third party, but the individual may be interviewed personally.

The SIB greatly exceeds its predecessors with respect to technical rigor. The norm group included 1,764 persons, selected to reflect the 1980 U.S. census with respect to gender, race, occupational level and status, geographic region, and community size. The norm tables were developed using technically advanced scaling techniques, and weighting procedures were used to match the norming sample with the census data.

The internal-consistency reliability of the SIB total scores equals or exceeds $r_{xx} = .96$ at all ages, indicating that the items of the scale are consistently measuring the same kinds of behaviors. Thus, with respect to overall reliability, the SIB is comparable to the most technically rigorous of the individually administered intelligence tests now available. The four adaptive-cluster scores of the SIB are considerably less reliable, however, and the individual subscale scores are not sufficiently reliable to be useful in decision-making.

Sufficient support for the validity of the SIB exists to indicate that the scale does, in fact, measure what it purports to measure. As would be expected theoretically, SIB scores increase with chronological age ($r = .68$ to $r = .82$), indicating that people become better adapted to their environments as they become more mature and gain greater experience. As would also be expected, SIB adaptive scores are, on average, lower for handicapped individuals than for nonhandicapped, and patterns of score deficit are consistent with expectancies based upon particular handicapping conditions.

Validity is further supported by evidence that SIB scores behave as would be predicted theoretically in relationship to other kinds of performance scores. For example, the SIB correlates strongly ($r = .59$ to $r = .91$) with the ABS-SE, another scale of adaptive behavior. Also as expected theoretically, SIB scores show considerable overlap with cognitive ability scores such as those for the Woodcock-Johnson Broad Cognitive

Ability (Woodcock, 1977). The maladaptive scores of the SIB also correlate satisfactorily with the Revised Behavior Problem Checklist (Quay & Peterson, 1983).

Vineland Adaptive Behavior Scales. The three editions of the Vineland Adaptive Behavior Scales (VABS; Sparrow, Balla, & Cicchetti, 1984) reflect an extensive redesign and restandardization of the Vineland Social Maturity Scale (Doll, 1965). Two of these editions, the Survey and the Expanded versions, are designated "interview" editions because they require interviews with a parent or caretaker. The third, the Classroom Edition, requires the individual's classroom teacher to complete a questionnaire rating form.

Four domains are common to all editions: (1) communication (receptive, expressive, and written language), (2) daily living skills (personal, domestic, and community skills), (3) socialization (interpersonal skills, play and leisure, and coping), and (5) motor skills (gross and fine motor skills). A fifth domain, maladaptive behavior, is included in the survey and expanded versions only. All editions of the VABS provide supplementary norm tables, which may be used to compare the individual being assessed to a population of similarly handicapped individuals (e.g., mentally retarded, emotionally disturbed, visually impaired, or hearing impaired individuals).

The Interview Editions. The two VABS interview editions are based on a norm group of 3,000 individuals, aged newborn to 18 years, 11-months, whose demographic characteristics closely matched those of the U.S. census population in 1980.

The reliability of the interview editions is well supported, indicating that these editions are consistently measuring the same kinds of behaviors, with regard to both overall performance (total score) and performance within each of several domains. Scores from the VABS Survey are accurate enough for most decision-making purposes, both at the composite-score level and at the domain-score level.

Even higher reliabilities are estimated (from data collected using the survey form) for the expanded form, which contains all the items in the survey form, plus an almost equal number of additional items.

The Classroom Edition. The classroom edition was normed on a sample of 1,984 children whose ages ranged from 3 years to 12 years, 11 months. This sam-

ple closely matched the U.S. population of 1980 in racial/ethnic makeup but tended to be somewhat weighted toward children from the north-central region of the United States, toward children growing up in urban communities, and toward children whose parents were college-educated. Like other versions of the VABS, the classroom edition demonstrates good-to-excellent reliability, both overall and within specific domains.

The validity of the VABS (i.e., its ability to measure what it purports to measure) is supported by several types of evidence: by positive correlations between VABS raw scores and chronological age, by performance differences on the VABS among the various supplementary norm groups, and by factor-analytic data supporting the underlying structure of the VABS. The validity of the VABS is further supported by evidence that people's VABS scores relate to their scores on other testing instruments in ways that would be predicted theoretically. VABS scores show moderate (positive) correlations with intelligence scores, for example.

In all editions, the VABS is generally regarded as a balanced, well-developed instrument with sound psychometric properties. Certain technical problems should be considered in using the VABS, however, especially when comparing an individual's long-term performance across several age levels, or when evaluating an individual's relative strengths and weaknesses among the four domains (Silverstein, 1986). Overall, the VABS is probably the most technically advanced adaptive-behavior scale currently available.

(See also: CULTURE AND INTELLIGENCE.)

BIBLIOGRAPHY

- BRUININKS, R. H., & MCGREW, K. S. (1987). *Exploring the structure of adaptive behavior* (Report Number 87-1). Minneapolis: University of Minnesota, Department of Educational Psychology.
- BRUININKS, R. H., WOODCOCK, R. W., WEATHERMAN, R. R., & HILL, B. K. (1984). *Scales of Independent Behavior* (SIB). Allen, TX: DLM Teaching Resources.
- DOLL, E. A. (1965). *Vineland Social Maturity Scale*, Circle

- Pines, MN: American Guidance Service. (Originally published in 1953.)
- HARRISON, P. L. (1987). Research with adaptive behavior scales. *Journal of Special Education*, 21, 37-68.
- HARRISON, P. L. (1989). Adaptive behavior: Research to practice. *Journal of School Psychology*, 27, 301-317.
- HEBER, R. (1961). *A manual on terminology and classification in mental retardation* (2nd ed.). Monograph supplement to the *American Journal of Mental Deficiency*.
- LAMBERT, N. (1981). *The AAMD Adaptive Behavior Scale-School Edition manual*. Monterey, CA: Publishers Test Service.
- LAMBERT, N., & WINDMILLER, M. (1981). *Diagnostic and technical manual, revised, AAMD Adaptive Behavior Scale-School Edition*. Monterey, CA: CTB/McGraw-Hill.
- LAMBERT, N., WINDMILLER, M., THARINGER, D., & COLE, L. (1981). *Administration and instructional planning manual, AAMD Adaptive Behavior Scale-School Edition*. Monterey, CA: CTB/McGraw-Hill.
- LUCKASSON, R. A., COULTER, D. L., POLLOWAY, E. A., REISS, S., SCHALOCK, R. L., SNELL, M. E., SPITALNIK, D. M., & STARK, J. A. (1992). *Mental retardation: Definition, classification, and systems of supports*. Washington, DC: American Association on Mental Retardation.
- MERCER, J. R. (1979). *System of multicultural pluralistic assessment technical manual*. San Antonio, TX: The Psychological Corporation.
- MERCER, J. R., & LEWIS, J. F. (1978). *System of multicultural pluralistic assessment*. San Antonio, TX: The Psychological Corporation.
- NIHIRA, K., FOSTER, R., SHELLHAAS, M., & LELAND, H. (1975). *AAMD adaptive behavior scale for children and adults, 1974 revision*. Washington, DC: American Association on Mental Deficiency.
- PLATT, L. O., KAMPHAUS, R. W., COLE, R. W., & SMITH, C. L. (1991). Relationship between adaptive behavior and intelligence: Additional evidence. *Psychological Reports*, 68, 139-145.
- QUAY, H. C., & PETERSON, D. R. (1983). *Revised Behavior Problem Checklist*. Coral Gables, FL: Authors.
- SILVERSTEIN, A. B. (1986). Nonstandard standard scores on the Vineland Adaptive Behavior Scales: A cautionary note. *American Journal of Mental Deficiency*, 91, 1-4.
- SPARROW, S. S., BALLA, D. A., & CICCETTI, D. V. (1984). *Vineland Adaptive Behavior Scales*. Circle Pines, MN: American Guidance Service.
- WOODCOCK, R. W. (1977). *Woodcock-Johnson Psycho-educational*

Battery: Technical report. Allen, TX: DLM Teaching Resources.

MARY P. DISIBIO

ADULTHOOD, ACHIEVEMENT IN How achievement varies across the adult lifespan has been studied formally for over 150 years (Simonton, 1991). Quetelet (1835/1968) applied statistical methods to analyze change in the output of successful plays over the careers of playwrights. Beard (1874) reported a study of relations between age and achievement in his book *Legal Responsibility in Old Age*. In the present century, interest in this topic was rekindled by Lehman (1953), whose studies of relations between age and achievement spanned thirty years, and by Dennis (1954, 1956, 1958).

Most of the work in this area has addressed two major issues. The first is the basic shape of the function that relates age and achievement, and how this shape can be explained. The second is whether this function is pretty much the same across disciplines and professions. The current state of knowledge about these issues will be summarized, and then some controversies, limitations, and remaining questions will be considered.

BASIC FUNCTIONAL RELATIONS BETWEEN AGE AND ACHIEVEMENT

Functional relations between age and achievement are expressed in age curves, which are plots of levels of achievement at various ages.

Typical Age Curves. The typical age curve is represented by a relatively rapid rise in productivity to a peak level of productivity, which occurs roughly at the end of the second decade of career experience, followed by a more gradual decline to perhaps about half the rate of productivity shown at peak performance. This view enjoys wide consensus, with two minor exceptions.

First, some disagreement exists about the existence of a single distinct peak. The two alternative proposals include a bimodal shape with two peaks, and a single, broad plateau, as opposed to a distinct peak (Simonton, 1988). With regard to the bimodal shape with

two peaks, the second peak commonly represents an apparent resurgence of productivity at retirement age or later (e.g., Haefele, 1962), perhaps as productive individuals contemplate their legacy (Simonton, 1989). This second peak, when it occurs, is smaller in magnitude than the peak that characterizes an individual's maximal level of performance. With regard to the existence of a broad plateau rather than a distinct peak, this possibility is somewhat difficult to evaluate definitively. Part of the difficulty is that whether a shape is characterized by a peak or a plateau is a bit subjective: One person's peak is another person's plateau. Another part of the difficulty is that a single study will typically have a heterogeneous sample of subjects drawn from different disciplines. If two disciplines each are characterized by a distinct peak, but the peaks occur at somewhat different points in experience, a combined sample made up of individuals from both disciplines will tend to show a broad plateau, even though each of the disciplines represented in the sample is characterized by a distinct peak (Simonton, 1988).

Second, some disagreement exists about the rate of decline in productivity in later years. The extent of decline varies, depending on how achievement is measured, with more decline apparent for measures of achievement that emphasize quality and less decline apparent for measures that emphasize quantity. For example, Lehman (1953) reported substantial decline in productivity with advancing age when only superior contributions were counted, but much less decline for contributions of lesser merit. Dennis (1956) found very little, if any, dropoff in productivity with advancing age, when the measure of productivity was total number of contributions without regard to quality. However, Simonton (1988a) has argued that these apparent differences in the shape of age curves as a function of quality of contributions may be artifacts of poor study design. For example, in some studies, a more selective sample was used in studies of major contributions compared to the samples used in studies of total output. Simonton (1988a) reported more similarity in age curves for major and lesser contributions when they were generated from a single sample of producers.

Explanations. Ignoring for the moment differences of opinion about the rate of decline in performance with aging, what explains the prototypic age

curve of increasing productivity to a maximal level of performance followed by a slower rate of decline? Three possible explanations will be reviewed.

Beard's (1874) Two-Factor Theory. According to this view, achievement depends on the levels of two factors: enthusiasm and experience. Enthusiasm provides the motivation for the long-term effort that is required for major achievement. Experience provides the ability to distinguish promising avenues from unpromising ones, and to capitalize on them. A balanced combination of considerable enthusiasm and experience is required for major accomplishment. An individual who is highly enthusiastic but lacking in experience is likely to jump about from one blind alley to the next. Conversely, an individual who is highly experienced but lacking in enthusiasm is likely to make mundane contributions.

The typical age-curve falls out of Beard's theory by assuming different developmental patterns for enthusiasm and experience. Enthusiasm is assumed to be at its peak early in a career and to decline thereafter, whereas experience is assumed to be limited early in a career and to increase thereafter. An optimal balance of enthusiasm and experience is achieved mid-career, and it is this optimal balance that accounts for the commonly observed peak in achievement.

Beard's theory appears to have a ring of truth about it, but its generality makes it difficult to come up with specific, testable hypotheses that could be used to refute the theory. The next explanation to be considered is similar to Beard's in spirit but provides more specificity for generating testable hypotheses.

Ericsson's Theory of Deliberate Practice. On the basis of examining the means by which experts achieve and maintain their elite levels of performance, Ericsson and colleagues have put forth a theoretical framework that can be regarded as a modern counterpart of Beard's two-factor theory (Ericsson, Krampe, & Tesch-Romer, 1993). According to this view, elite levels of adult performance are the end result of a prolonged, deliberate effort to improve performance, while negotiating motivational and external constraints.

The kind of deliberate practice that Ericsson and colleagues refer to is intense, effortful, and focused on the goal of improving current level of performance. This kind of deliberate practice can be distinguished from the routine practice that is an automatic by-

product of experience, and that does not routinely result in elite levels of performance. It has long been known that adults typically perform at a level well below their maximum possible level, even for tasks that are done repetitively for years (Thorndike, 1921). Thus, adults write more slowly and less legibly than they are capable of, and clerks with many years of experience add numbers much more slowly than they are capable of doing.

Across a wide variety of disciplines and domains, top achievers are individuals who have engaged in a decade or more of deliberate practice that has been carried out for the purpose of improving performance. Despite the common belief that highly talented individuals acquire exceptional levels of performance with little time and effort, careful examination of the histories of so-called prodigies fails to find substantial evidence for the attainment of elite levels of performance without at least a decade of serious training. In chess, for example, no one has achieved the level of international grandmaster with less than about a decade of intense preparation (Simon & Chase, 1973). Bobby Fischer and Judit Polgar are regarded as the best examples of child prodigies, having attained international grandmaster status at the earliest recorded age of 15. However, Fischer learned the game and studied his first book of chess games at age 6. At age 7, Fischer was tutored by the president of the Brooklyn Chess Club, and at age 12, he joined the Manhattan Chess Club, which was among the strongest in the world, profiting from interest shown by chess expert Jack Collins. Thus, even someone with the talent of Bobby Fischer was at best a year shy of the required decade of training. Similarly, Judit Polgar was tutored and coached by her father from age 4 or 5. The requirement of a decade of training for elite levels of performance has been observed in a wide variety of domains, including music, mathematics, tennis, swimming, running, medical diagnosis, and even livestock evaluation (Ericsson, Krampe, & Tesch-Romer, 1993).

The decade of intense, deliberate study and practice requires enormous and sustained motivation, as well as environmental circumstances that support such effort. The best predictor of ultimate level of achievement is individual differences in the cumulative amount of deliberate practice. When this framework is applied to the typical age-curve, the increasing level

of productivity reflects the accumulation of deliberate practice. A tendency for a decrease in performance with advancing years may reflect a reduction in effortful activity as a consequence of such factors as motivational burnout, the elimination of major extrinsic motivational factors for researchers, such as the attainment of tenure and of the highest academic rank of professor, or competing priorities, such as family responsibilities, and in some cases, problems associated with declining health (Simonton, 1977).

Changes in Intellectual Functioning: Horn-Cattell Model.

The distinction between two kinds of intelligence—fluid and crystallized—is the most relevant theory of changes in intellectual functioning across the adult life span for understanding the relation between age and achievement. Fluid intelligence, which corresponds to basic, abstract reasoning ability and is closely linked to physiological and neurological functioning, appears to decline with aging. Crystallized intelligence, which corresponds to cultural knowledge and skills, increases or remains stable over much of the adult life span. (See FLUID AND CRYSTALLIZED INTELLIGENCE, THEORY OF.)

The evidence concerning changes in fluid and crystallized intelligence is difficult to interpret because many studies suffer from problems associated with cohort differences. *Cohort differences* refer to the fact that different generations of individuals have different educational and intellectual experiences due to changes in culture. This creates problems for studies that compare young adults with older adults, in what is known as a cross-sectional design, because the young adults typically have been exposed to a richer intellectual environment than have the older adults. Cohort differences can produce spurious decline in intellectual functioning, or can make real decline appear to be greater than it actually is. Studies that do not suffer from cohort-difference problems suggest that some decline exists on average for most kinds of tasks found on IQ tests, although (1) the decrements up through age 60 are too small to be of practical importance; (2) decrements from ages 60 to 80 appear to be more problematical; and (3) the extent of decrement varies across individuals, with some adults showing no decrement through age 70 (Schaie, 1983; Schaie & Hertzog, 1983).

The upshot is that whereas there appears to be some decline in fluid intelligence with aging, well-

practiced and skilled performances remain unaffected and can even show continued growth. For most individuals, sufficient reserves in fluid intelligence exist so that the modest decline associated with aging is not a serious limitation for most kinds of intellectual tasks (Dixon & Baltes, 1986).

Simonton's Chance-Configuration Theory. Simonton (1988a) has proposed a model in which age-curves are a function of a mathematical equation in three parameters: creative potential, or the total number of new ideas an individual is capable of producing over a lifetime; ideation rate, the rate at which new ideas are generated; and elaboration rate, the rate at which ideas are converted into finished products. Thus, age-curves reflect the rate by which creative potential is converted into new ideas and elaborated into finished products. Individual differences in career trajectories can be traced to individual differences in each of these attributes.

Explaining the specific workings of the model requires more mathematics than is appropriate for the present context, but the bottom line is that the model generates age-curves that closely follow observed age-curves when appropriate values are used.

DIFFERENCES ACROSS DISCIPLINES

How comparable are age-curves across disciplines? Does the prototypic shape of productivity advancing to a peak and then showing modest decline thereafter apply to different disciplines such as mathematics, psychology, and medicine? If the overall shape is comparable across disciplines, how comparable are such things as location of peak performance and rates of increasing and decreasing performance?

In general, the overall shape of age-curves is similar across disciplines, although they vary in the sharpness of the peak and the corresponding rates of rise and decline in productivity. Simonton (1991) carried out a biographical study of 2,026 scientists and inventors who were notable enough to be mentioned in one of three selective biographical dictionaries. No member of the sample was still living. The earliest member was born in 1450; the average birth year was 1790. Nine scientific disciplines were represented: mathematics, astronomy, physics, chemistry, biology, medicine,

technology, earth sciences, and other sciences. Modest but reliable differences were found in average age when the scientists produced their best work, ranging from mathematicians, with an average age of 38.8, to earth scientists, with an average age of 42.5.

In general, disciplines with relatively early and sharp peaks, and with considerable decline thereafter, include theoretical physics, pure mathematics, and lyric poetry. Disciplines with relatively later and modest peaks, and with little decline evident, include novel writing, history, philosophy, and medicine (Simonton, 1988).

CONTROVERSIES, LIMITATIONS, AND REMAINING QUESTIONS

As in any research endeavor, the area of achievement in adulthood is characterized by controversies, limitations, and remaining questions. A few of the most salient examples will be mentioned.

Given how difficult it is for organizations to assess the performance of their executives, for universities and professional organizations to assess the contributions of scientists, and for any consensus to be reached concerning the merit of artistic productions, it should come as little surprise that one controversy that has dogged this area of inquiry from the beginning is how to assess achievement. If three researchers set about to study the relation between age and outstanding achievement in a field such as theoretical physics, and each investigator began with completely different lists of the supposed top twenty-five contributors to the field, it should not be surprising if there is little overlap in their results and conclusions.

An example of a controversy about how to assess achievement is the issue of quality versus quantity. As mentioned previously, Lehman's work (1953) emphasized selectivity in the works and the contributors that he included in his samples, and his results suggest gradual yet substantial decline in achievement with advancing age. Conversely, Dennis (1956) emphasized total productive output, as opposed to counting only works judged to be major on some criterion, and his results suggest much less and perhaps even no decline in achievement with advancing age. Which approach to measuring achievement is the more appropriate is

unclear, as is the extent to which results are affected by whether an investigator emphasizes quality or quantity when measuring achievement. The existing data suggest a pattern of finding a less distinctive peak and less subsequent dropoff in achievement as the emphasis shifts from quality to quantity of productions. However, Simonton's "constant probability of success model," in which the probability of having a major, creative "hit" is a constant probability of the number of "times at bat" or total number of works produced, would suggest that similar results should emerge whether one emphasizes quality or quantity, provided sufficient samples of works are incorporated in the analysis (1988, 1988a).

In scientific disciplines, perhaps the most widely agreed-upon indicator of the importance of a work is a citation count, which is a count of the number of times a target work is cited or mentioned in published articles. Yet the process by which some articles become more widely cited than others is complex and largely unstudied (Shadish, 1989), and it would be naive to believe that citation count is a direct measure of quality or importance. A 1951 article by O. H. Lowry on protein measurement was cited 50,000 times between the years 1961 and 1975, a count that is orders of magnitude more than that for Einstein's article on his unified field theory. Few scientists, including Lowry himself, would suggest that Lowry's contribution was more important than Einstein's contribution (Garfield, 1979). Lowry's contribution was to describe a method for measuring proteins that was, in his own words, "... a trifle better or easier or more sensitive than other methods, and of course nearly everyone measures protein these days" (cited in Garfield, 1979). Citation counts favor methods such as Lowry's over theoretical contributions such as Einstein's. For a current example of this bias, see the list of the ten most highly cited articles assembled in a 1992 centennial celebration of the American Psychological Association, published in the journal *Psychological Bulletin* (November, 1992). Seven of the ten most highly cited articles were methodological in nature.

Perhaps the most salient limitation of our knowledge about the relation between age and achievement is a fundamental lack of data. Much of the data represent a cross-sectional assemblage (i.e., works from

individuals differing in age are collected in a single measurement), as opposed to a longitudinal design in which a group of individuals is followed over their careers, which confounds true developmental change with artifactual change attributable to societal differences experienced by different-aged cohorts or groups. Many of the samples represent a heterogeneous mix of small numbers of individuals drawn from diverse disciplines and domains, as opposed to large samples drawn from within a domain, which can produce results that are not characteristic of any of the domains represented in the heterogeneous sample.

The reason that first-rate longitudinal studies involving large numbers of subjects, measures, and years are not the norm is simple: They are extremely difficult to do. (A newly hired assistant professor who devotes the majority of her time to a decade-long longitudinal study of age and achievement will find herself drummed out by her university when it is time to decide on whether to give her tenure, which routinely will be several years before her study is complete.) The longitudinal studies that are done are likely to be characterized by a limited set of dated measures. It is impractical to give large numbers of subjects large numbers of measures over an appreciable period of time unless one has access to an extraordinarily large and stable source of funding. Decisions about what measures to obtain, which must be made at the beginning of the study, rarely are ideal in hindsight upon completion of the study a decade or more later.

In addition, much of the data are not contemporary, which makes one wonder how successfully the results can be applied to modern society, given the phenomenal changes that have occurred even in the last half of the present century. For example, the previously mentioned study of differences in relations between age and achievement across disciplines (Simonton, 1991), perhaps the best of its kind, had a sample of scientists whose average year of birth was 1790! An associated limitation is that by their very nature, the design of most studies is retrospective (i.e., they are historical studies) as opposed to the more powerful prospective design (of, for example, predictive studies).

Given the existing literature, the most salient remaining questions appear to center on the nature of

the developmental and sociological mechanisms that underlie change in achievement associated with aging. The bulk of existing knowledge consists of empirical facts—average age-functions for various disciplines, for example. Some progress has been made in identifying and evaluating possible explanations of the empirical facts, but progress to this point has been limited, and far removed from the long-term goal of describing underlying mechanisms. Given the extraordinary challenge offered by this subject matter, this present state of knowledge is nothing short of remarkable, despite being far closer to providing the beginning of the story than the end.

(See also: JOB PERFORMANCE; OCCUPATIONS.)

BIBLIOGRAPHY

- BEARD, G. M. (1874). *Legal responsibility in old age*. New York: Russell Sage Foundation.
- COLE, S. (1979). Age and scientific performance. *American Journal of Sociology*, 84, 958–977.
- DENNIS, W. (1954). Productivity among American psychologists. *American Psychologist*, 9, 191–194.
- DENNIS, W. (1956). Age and productivity among scientists. *Science*, 123, 724–725.
- DENNIS, W. (1958). The age decrement in outstanding scientific contributions: Fact or artifact? *American Psychologist*, 13, 457–460.
- DIXON, R. A., & BALTES, P. B. (1986). Toward life-span research on the functions and pragmatics of intelligence. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence* (pp. 203–235). New York: Cambridge University Press.
- ERICSSON, K. A., KRAMPE, R. T., & TESCH-ROMER, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review*, 100, 363–406.
- GARFIELD, E. (1979). Is citation analysis a legitimate evaluation tool? *Scientometrics*, 1, 359–375.
- HAEFELE, J. W. (1962). *Creativity and innovation*. New York: Reinhold.
- HARGENS, L. L., McCANN, J. C., & RESKIN, B. F. (1978). Productivity and reproductivity: Fertility and professional achievement among research scientists. *Social Forces*, 57, 154–163.
- HORN, J. L. (1980). Concepts of intellect in relation to learning and adult development. *Intelligence*, 4, 285–317.
- HORN, J. L. (1982). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 237–278). New York: Plenum.
- HORN, J. L., & CATTELL, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, 57, 253–270.
- HORN, J. L., & CATTELL, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107–129.
- HORNER, K. L., RUSHTON, J. P., & VERNON, P. A. (1986). Relation between aging and research productivity. *Psychology and Aging*, 1, 319–324.
- LEHMAN, H. C. (1953). *Age and achievement*. Princeton, NJ: Princeton University Press.
- MCDOWELL, J. M. (1982). Obsolescence of knowledge and career publication profiles: Some evidence of difference among fields in costs of interrupted careers. *American Economic Review*, 72, 752–768.
- QUETELET, A. (1835/1968). *A treatise on man and the development of his faculties*. New York: Franklin.
- SCHAIK, K. W. (1983). The Seattle longitudinal study: A 21-year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 64–135). New York: Guilford Press.
- SCHAIK, K. W., & HERTZOG, C. (1983). Fourteen-year cohort-sequential analysis of adult intellectual development. *Developmental Psychology*, 19, 531–543.
- SHADISH, W. R. (1989). The perception and evaluation of quality in science. In B. Gholsen, W. Shadish, R. Neimeyer, & A. Houts (Eds.), *Psychology of science: Contributions to metascience* (pp. 383–423). New York: Cambridge University Press.
- SIMON H. A., & CHASE, W. G. (1973). Skill in chess. *American Scientist*, 61, 394–403.
- SIMONTON, D. K. (1977). Creative productivity, age, and stress: A biographical time-series analysis of ten classical composers. *Journal of Personality and Social Psychology*, 35, 791–804.
- SIMONTON, D. K. (1988). Age and outstanding achievement: What do we know after a century of research? *Psychological Bulletin*, 104, 251–267.

- SIMONTON, D. K. (1988a). *Scientific genius: A psychology of science*. New York: Cambridge University Press.
- SIMONTON, D. K. (1989). The swan-song phenomenon: Last-works effects for 172 classical composers. *Psychology and Aging*, 4, 42–47.
- SIMONTON, D. K. (1991). Career landmarks in science: Individual differences and interdisciplinary contrasts. *Developmental Psychology*, 27, 119–130.
- THORNDIKE, E. L. (1921). *The psychology of learning* (Vol. 2). New York: Teachers College Press.

RICHARD WAGNER

AFRICAN AMERICANS Since the early twentieth century an acrimonious controversy has focused on the origins of a persistent and irrefutable difference between the test performances of African Americans and white Americans (Brody, 1992; Osborne & McGurk, 1982). For some persons, the distinction represents an inborn difference between the populations. For others, the distinction is more accurately explained by environmental factors. Rather than debating these contentions, this article will discuss the interaction of the genetic and environmental positions. Robert Sternberg (1988, p. 69) has posited that “intelligence is mental activity underlying purposive adaptation to, shaping of, and selection of real-world environments relevant to one’s life.” According to him, intelligent behavior can be understood within a framework that depicts adaptation as a function of an external problem context, an internal problem-solving mechanism, and a procedure for gauging the problem solution.

Evidence strongly suggests that the historical circumstances of African-American families have resulted in adaptive strategies that form culturally distinct ways of coping and learning. Research has described three patterns that typically characterize African Americans’ approach to problems as (1) a collective and relational process (Young, 1970); (2) occurring within a highly stimulating and active environment (Allen & Boykin, 1991; Boykin, 1979, 1983); and (3) expressed in a highly individualistic and “unique style” (Boykin, 1979; Hale-Benson, 1986). This article will examine the context of African-American family life as the source of

an enduring and distinctive problem-resolving strategy and learning style.

THE CONTEXT OF FAMILY LIFE

Generally speaking, families offer opportunities for crucial affective and instrumental support for their members, providing the context for physical maintenance, familial affection, and social control. Gary Lee (1977) argued that although the universal presence of family does not imply a universal structure, the family often includes an association of adults of both sexes and dependent children. Indeed, a central aspect of family life involves the bearing, rearing, and caring of children.

A family is influenced and facilitated by the opportunities and constraints of its social context (Bronfenbrenner, 1977; Myers, 1982; Wicker, 1979; M. N. Wilson, 1986). To ensure its own continued existence, a family adapts available resources to normal and non-normal crisis-stress events (Barbarin, 1983). Family resources involve the ability of family members to contribute tangible help, such as material support, income, child care, and household maintenance, and nontangible aid, such as expressive interaction, emotional support, instruction, and social training and regulation. Marriage, fecundity, and death typify normal stress events, whereas hospitalization and unemployment are nonnormal stress events (Carter & McGoldrick, 1988; Duvall, 1971). A common stressful situation in the African-American community is the lack of adequate adult resources in single-parent family units.

Studies have suggested that single-parent family units are a primary reason for the formation of extended-family support networks in the African-American community (M. N. Wilson, 1986, 1989). That is, an extended family is formed when a viable family unit—a nuclear family, a childless couple, or a single parent—absorbs either an orphaned child, a dependent adult, or a vulnerable single-parent family unit. Once formed, the extended family is multigenerational and extranuclear, and occupies the most of the family life span.

African-American Family Life. The African-American family differs from the white family in a

number of ways. Whereas the white family structure historically has implied opposite-sex parents living with their children in one household, the African-American family structure is more often extended, not limited to membership in a particular nuclear family, household, or set of blood relatives (Glenn, 1992; Gutman, 1976; Martin & Martin, 1978; H. P. McAdoo, 1991; Oliver, 1988). The African-American extended family characteristically has a familial interaction network involving relatives, friends, and neighbors who provide emotional and economic support (Cazenave & Smith, 1990; H. P. McAdoo, 1991; Slaughter & Dilworth-Anderson, 1988). This system serves as a buffer against negative ecological forces and provides a coping response for external stress. Harriette McAdoo (1991, 1992) suggested that extended families provide an alternative structure of support that supplements the necessities of nuclear family units and protects the integrity of the African-American family. For example, nurturance, material assistance, and mutual aid are provided by a network of kin and friends to dependent loved ones (Beck & Beck, 1984, 1989; Martin & Martin, 1978).

Researchers of social networks of African Americans have consistently documented frequent socializing among African-American family members, a high degree of residential propensity among related households, and an emphasis on participation in family occasions, especially funerals, holiday celebrations, and birthdays (Aoyagi, 1978; Aschenbrenner, 1978; Hale-Benson, 1986; Martin & Martin, 1978; H. McAdoo, 1991; McLanahan, Wedemeyer, & Adelberg, 1981; Stack, 1974; M. N. Wilson, 1984). African-American women often list relatives as friendship contacts (Hale-Benson, 1986; Martin & Martin, 1978; Stack, 1974). Also, African-American adults rely on family members for advice and guidance (Hatchett, Cochran, & Jackson, 1991; Stack, 1974). Family members often turn to their elders for counsel on job choices, major purchases (Landry & Jendrek, 1978; Martin & Martin, 1978), and decisions concerning children (Martin & Martin, 1978; H. McAdoo, 1978; Stack, 1974). Therefore, the active, vibrant, and collective nature of extended-family support and interactions reflects an important aspect of African-American family life (Aschenbrenner, 1978).

Historical Development of the African-American Family. Among African Americans, as among other American ethnic groups, the development of the family was affected by its American context. It is marked by a tumultuous history totally unlike that of other ethnic groups, one that involved involuntary migration to the United States and then a protracted period of enslavement and, after emancipation, segregation and inferior status based on race (Hacker, 1991; Lieberman, 1980; Perlmann, 1988). African-American family development can be traced to the development of African-American slave families. Many researchers believe that an African-American familial tradition existed throughout the period of slavery. Others have recognized the similarities in family structure and values between African Americans and Africans (H. P. McAdoo, 1992; Young, 1974).

Historical research suggests the existence of strong familial associations among slaves, even when the rules of slavery did not sanction such relationships. Because of the disruption of African families and kinship patterns caused by the slave trade, slaves invested nonkin with symbolic kin status (Genovese, 1976; Gutman, 1976). As a result, African-American slave family and related kin groups emerged from the initial disruptions associated with enslavement. The development of multigenerational linkages among slave families was accompanied by an understanding of family and kin obligations. Several findings, confirmed by replicated historical studies, suggest that 70 percent of the children born into slavery were born into long-standing conjugal relationships (Agresti, 1978; Gutman, 1976; Otto & Burns, 1983); family units were eventually reconstituted or blended into stepfamily situations when families were separated by one member being sold to another plantation (Dykstra & Manfra, 1985; Meacham, 1984); and plantations usually contained large extended, multigenerational, and collateral families by 1825 (M. Cohen, 1984; Genovese, 1976; Herskovits, 1966). The institution of the African-American family was not destroyed during slavery; it existed as a mechanism for coping and survival (Agresti, 1978; Fogel & Engelman, 1974; Genovese, 1976; Gutman, 1976; Herskovits, 1966; Meacham, 1983).

Like slavery and emancipation, the massive rural-to-urban migration that began during the early 1920s

and continued until the mid-1950s (Franklin, 1967) represented an era of intense pressure on the African-American family. Several researchers (Aoyagi, 1978; Flanagan, 1978; Martin & Martin, 1978) have suggested that the transition to urban life did not deter the development of the African-American family; on the contrary, the extended-family network played a pivotal role in the rural exodus. According to William Flanagan (1978), extended families sponsored the initial migration of their family members. For instance, a family member would be sent to obtain work in a city to earn money for the family or a rural family member would join the family urban kin. After some time, whole families would reside in a particular urban area. Next, the families established a social-welfare system, which allowed for the care of dependent children and adults. Because obtaining a job was difficult even in the cities, many adults temporarily relied on the extended family for food, care, and shelter. Finally, a channel of urban income distribution developed; less-fortunate family members received subsidies from others in the kin network.

Current Status of the African-American Family. During the last three decades of the twentieth century, the effects of dismal socioeconomic realities and unstable interpersonal relationships led to low incomes and high amounts of changes in household structures for increasing numbers of African-American families (Angel & Tienda, 1982; Cutright, 1971; Hofferth, 1984; McLanahan, Astone, & Marks, 1991). Whereas most African-American families are not poor, a higher proportion of them has always lived below poverty levels than have white families: Since 1960, the average rate for African-American families has been 3.8 times higher than the proportion of white families living below poverty levels (Duncan, 1968; Lieberman, 1980; Reid, 1982; Wacquant & Wilson, 1989). Among one-parent families, poverty is particularly evident; 62 percent of African-American one-parent families are poor.

The current status of poor African Americans is greatly influenced by changing social and economic structures, which have strained the resources and ability of the family system to respond. Changes in the nature of work and technological advances have altered the nature of American poverty, influenced Af-

frican Americans' ability and opportunity to find work, and adversely affected the African-American family (Hill, 1990; Hochschild, 1989; Nathan, 1989; Reid, 1982; W. J. Wilson, 1991).

The rate of poverty among African Americans correlates highly with family size and family composition (Allen, 1979; Angel & Tienda, 1982; Hill, 1990; Hofferth, 1984). Walter Allen (1979) and Sandra Hofferth (1984) found that when race and marital status were adequately controlled, socioeconomic status was highly predictive of family formation in both African-American and white families. Although differences were small, the data supported the greater likelihood that African-American families used residential sharing as a way of reducing the effects of low income. Comparing the propensity to form extended families among African-American, Hispanic, and white American families, researchers demonstrated that African-American family formation was also reflected as a group-specific difference. Marta Tienda and Ronald Angel (1982) found that African and Hispanic American households were more likely to share residence with extranuclear members who were contributing to the overall household income. In white households, extranuclear family members did not significantly contribute to the generation of household income; on the contrary, they were more often the beneficiaries of household income. Finally, Angel and Tienda noted that single mothers and their children, who had the highest likelihood of living below the poverty level, were the most likely to be extended (Angel & Tienda, 1982; Tienda & Angel, 1982).

Moreover, several researchers have suggested that race-specific effects exist not only at poverty level but at each socioeconomic level. For example, African-American families have consistently had lower incomes than whites. The consumable income (family income per family member) of working- and middle-class African-American families has typically been affected by their tendency to have more child and adult household members than white families. Although increasing numbers of African-Americans have attained favorable socioeconomic levels, their adult children's ability to maintain a comparable socioeconomic level is less than that of their white counterparts (Cutright, 1971; Duncan, 1968; Glick & Norton, 1979; Hacker,

1991; Hill, 1990; Lieberman, 1980; Lieberman & Carter, 1979; H. P. McAdoo, 1978, 1991; Perlmann, 1988; Reid, 1982).

Frequent family changes and high numbers of marital disruptions suggest that African-American children experience alarmingly high amounts of fluctuation in their living situations (Furstenberg, Brooks-Gunn, & Morgan, 1987; Slesinger, 1980). For example, 64 percent of African-American parents were divorced, separated, widowed, or never married; 63 percent of African-American births were to unmarried mothers; and 30 percent of African-American births were to adolescents (U.S. Bureau of the Census, 1993).

For an African-American child, a nearly equal likelihood exists that the child will live in a single-parent family (42% of African-American children) as in a dual-parent family (40% of African-American children). The remaining percentage of African-American children live with a parent and another relative (18% of African-American children; U.S. Bureau of the Census, 1993). In fact, three times as many African-American children under age 18 are living with a grandparent as are white children.

Single mothers, including those who are divorced, separated, widowed, or never married, account for 94 percent of African-American single parents (McLanahan & Garfinkel, 1989; Reid, 1982). A single mother's living arrangement is affected by the age of her child and her level of educational attainment (Colletta, 1979, 1981). Typically, a single-parent family unit involving a young mother of limited education having few children and low income will likely share a residence with extended-family members. Consequently, the most common composition of extended households involves a single-parent family structure that includes a mother, her children, and the single mother's mother.

Although men are conspicuously absent in many African-American families, the role of father-husband is very important in the African-American family (Cazenave, 1979; J. L. McAdoo, 1993; Scanzoni & Scanzoni, 1981; Slaughter & Dilworth-Anderson, 1988). After the mother, the adult family members most likely to be involved in child care are the father and the grandmother, respectively (Christman, 1990; Danziger & Radin, 1990; Slaughter & Dilworth-Anderson, 1988; M. N. Wilson et al., 1990). African-American

fathers do not have a history of exclusive responsibility for a family's economic situation. Mothers have always been significant participants in the work force and major contributors to the family's income. African-American families have depended on the wages of both-parents. Furthermore, African-American fathers have a long history of participating in the nurture and socialization of children. Fathers' interactions with their children may reflect not only role flexibility in the family but also fathers' desire to have a significant part in the development of their children. Moreover, African-American women expected their men to provide both material and emotional support to their children (Leslie & Anderson, 1988; J. L. McAdoo, 1988).

A significant proportion of African-American men experience chronic difficulty obtaining and maintaining stable and legitimate employment that can adequately support a family (Testa et al., 1989; W. J. Wilson, 1991; Wojtkiewicz, McLanahan, & Garfinkel, 1990). It is the employment and economic instability of African-American men that causes a considerable amount of family stress and vulnerability. Moreover, although divorce, desertion, separation, and extramarital births reflect significant reasons for father-husband absence, joblessness, incarceration, and mortality are other important factors that influence the high disparity in the male-female ratio (Darity & Myers, 1983, 1984).

ENVIRONMENTAL INFLUENCE ON ACHIEVEMENT

Confronted by the high incidence of social, economic, and interpersonal stresses and crises in their community, low-income and working-class African-American families have frequently used coping strategies and practical resolutions that reflect an active, affiliative, and flexible approach to environmental stresses (Garbarino et al., 1992). In particular, successful families are able to sustain some semblance of stable, supportive, and emotional relationships (although not necessarily traditional ones); create an open, reassuring, and stimulating climate; and include adult models of functional survival skills that encourage constructive coping and support. Historical and cultural aspects of African-American family life reinforce the value of the extended-family system and reliance on a

collective and relational approach that is both vibrant and original and that contributes to the ethnic and social milieu of African-American children's achievement (Hale-Benson, 1986; Moore, 1987).

In general, the family plays a primary role in the development and socialization of the child's cognitive abilities. Researchers have suggested that socialization usually occurs as a function of four basic child-rearing behaviors and activities: nurture, which refers to parental support, affection, and/or encouragement; demand, which refers to parental actions involving supporting achievement-oriented behavior or skill acquisition; control, which refers to parental actions involving reinforcing, setting, and enforcing limits on socially appropriate behavior; and punishment, which refers to parental punitive actions, such as the withdrawal of affection and/or privileges (Baumrind, 1971, 1972; Becker, 1964; Belsky, Robins, & Gamble, 1984; Emmerich, 1977; Maccoby, 1980; J. L. McAdoo, 1993).

Researchers have determined that low-income and working-class African-American parents focus most of their attentions on a combination of the four child-rearing activities to produce an atmosphere for encouraging individualistic and independent behaviors, demanding premature autonomous behaviors and maturity, maintaining strict authority and/or enforcing behavior and role correctness, and applying arbitrary rules (Bartz & Levine, 1978; Durrett, O'Bryant, & Pennebaker, 1975; Hale-Benson, 1986; Kelley, Power, & Wimbush, 1992; Kelley, Sanchez-Hucles, & Walker, 1993; J. L. McAdoo, 1988; M. N. Wilson, Hinton, et al., 1990, 1992). Michelle Kelley and her colleagues (1992, 1993) observed that strict disciplinary practices were associated with a mother's education and age, a father's absence, religious beliefs, and the parents' concerns for the child's safety and fears of child victimization.

The specific influences of the extended family are probably more indirect than direct in nature. For example, the effects of the extended-family involvement usually take the form of relieving the single mother of some household tasks, but not primary child-care tasks; increasing the opportunity of adult-adult exchanges, but not nonmaternal adult-child exchanges; and providing emotional support to the mother (M. N. Wilson, Kohn, et al., 1992). Nevertheless, research has

suggested that the participation of nonmaternal adults does not significantly lessen a mother's child-care and household responsibilities (Hurlbert, 1990; Slaughter & Dilworth-Anderson, 1988; M. N. Wilson, Hinton, et al., 1990; M. N. Wilson, Tolson, et al., 1990).

Overall, children's levels of educational achievement and social adjustments are positively affected by various living arrangements of extended family structures. Children from two-parent, mother-grandmother, mother-aunt, and mother-other families achieved and adjusted at adequate rates (Kellam et al., 1977, 1982). Children from mother-only and mother-stepfather families, however, were functioning below the rates of the other children. Still, research suggests that child care functions best as a cooperative venture involving the availability of at least two adults (Broman, 1988; Pearson et al., 1990; Tolson & Wilson, 1990; Wilson & Tolson, 1986).

Elsie Moore (1985, 1986, 1987) argued that the ethnicity of the rearing environment, not just the socioeconomic status and maternal education level, exerts a significant influence on children's styles of responding to standardized intelligence and achievement tests. Her studies have consistently indicated that the average test performances of African-American children who were adopted by white parents exceeded the test performances of African-American children who were adopted by African-American parents. In particular, Moore suggests that African-American families may emphasize a social, as opposed to an object, orientation; affective and emotional, as opposed to calm and reserved, modes of communication; and tolerance for, as opposed to susceptibility to, varying levels of sensory stimulation (Moore, 1987).

A salient aspect of low-income and working-class African-American households is the presence of high levels of environmental and social stimulation. High environmental and social stimulation is indicated by high numbers of people living in the home, frequent visitors in the home, and high amounts of lively social interactions (Boykin, 1983). According to A. Wade Boykin, children living in households with high levels of social and environmental stimulation are likely to develop high activity levels.

Additionally, Brenda Allen and Boykin (1988) examined children's activity levels and the level of stimulation in the homes of working-class African-

American children. The study revealed a positive correlation between the reported level of sensate stimulation in the home and a child's activity level. When comparing African-American and white children, Allen and Boykin (1991) found that African-American children's performances on a word-acquisition task was superior to the performances of white children during a high music stimulation trial, but white children surpassed African-American children during the low music stimulation trial. Furthermore, Boykin (1979) pointed to considerable anecdotal evidence and observations of teachers to support the claims of an increased behavioral vibrancy among African-American children, especially those from lower- and working-class families.

Culturally distinct patterns of environmental and social stimulation may partially account for the at-risk educational status of many low-income and working-class African-American children. Several studies have demonstrated that environmental factors mediate the relationship between parent and child intelligence scores (Bradley, Caldwell, & Rock, 1992; Elardo, Bradley, & Caldwell, 1975; Scarr & McCartney, 1983). Specifically, Robert Bradley and his associates (1993) reported that the home environment revealed minimal mediation effect for scores at year 1 but significantly stronger effects for year 3. Bradley and his associates used a simple model, however, that did not account for other exogenous factors, such as family composition, crowding, and paternal IQ.

Indeed, Arnold Sameroff and others (1993) have suggested that although statistically significant outcomes are associated with a single-risk factor, these differences rarely explain a large proportion of the outcome variance. Their research revealed that the cumulative environmental-risk indices included the family's minority status, parents' occupation, mother's education, family size, family support, life events, parenting perspectives, anxiety, and mother's mental-health status. Sameroff and his associates identified multiple-risk factors that influence the stability of intelligence scores from preschool to adolescence. These factors are independent of mother's intelligence quotient (genetic) and socioeconomic status (social) factors.

Other researchers have argued that special programs can alter early adverse environmental effects on

childhood achievement levels. Specifically, a program of intensive, systematic early intervention should precede continuous educational intervention over the entire course of the elementary- and middle-school grades. Intensive programs involve daily encouragement of exploration, teaching basic skills, celebration of developmental advances, guided rehearsal and extension of new skills, and protection from inappropriate disapproval, teasing, or punishment (Ramey & Campbell, 1987; Ramey & Ramey, 1992).

Attaining educational opportunities represents a critical challenge confronting low-income and working-class African Americans. Although the high school completion rate for low-income and working-class African Americans increased from 10 percent in 1940 to 73 percent in 1990 (Hill, 1990; Reid, 1982), technological changes often required postsecondary education. Low-income and working-class African-American parents often stress the necessity of their children acquiring advanced education and training in order to escape economic disadvantages (Hochschild, 1989; Nathan, 1989; Reid, 1982; Wacquant & Wilson, 1989). Many low-income African-American parents continue to believe that their children's social and economic conditions will continue to improve as long as they have a chance for advanced education. Because of limited information and unfamiliarity about career alternatives, however, many low-income and working-class African-American parents have unrealistic educational aspirations for their children.

CONCLUSION

The African-American extended family organization clearly functions as a coping mechanism that is characterized as a collective or relational process (Martin & Martin, 1978; Young, 1970; Zollar, 1985; Zollar & Honnold, 1988) occurring within a highly stimulating, vibrant environment (Boykin, 1979, 1984; Ogbu, 1988) and filled with highly individualistic and unique forms of expressions (R. Cohen, 1969; Hale-Benson, 1986; Rutledge, 1988; Young, 1970). The extended familial support system is activated during family crises and dire social circumstances, such as poverty, unemployment, extramarital births, and marital dissolutions. It is easy to fathom that child-rearing and socialization bolstered by the daily participation in a socially active,

vibrant, and flexible system would likewise produce children who invariably display manifestations of such a lively familial background. Moreover, such a problem-solving approach may act to facilitate and/or impede aspects of a child's ability to become involved in alternative activities or learning situations (R. Cohen, 1969).

Indeed, the experiential background of the low-income and working-class African-American children may differ considerably from the experiences of white children (Samuda, 1974). Consequently, a serious detriment may occur to those who are evaluated by means of standards that are alien to them. Hence, the legitimate concerns of African-American parents are not focused on the consistent differences in test performances per se but on the resulting policies that may restrict their children's chances to obtain the social and economic benefits of the United States.

(See also: ETHNICITY, RACE, AND MEASURED INTELLIGENCE; RACE AND INTELLIGENCE; RACE AND IQ SCORES.)

BIBLIOGRAPHY

- AGRESTI, B. F. (1978). The first decades of freedom: Black families in a southern county, 1870–1885. *Journal of Marriage and the Family*, 46, 697–706.
- ALLEN, W. R. (1979). Class, culture, and family organization: The effects of class and race on family structure in urban America. *Journal of Comparative Family Studies*, 10, 301–313.
- ALLEN, B. A., & BOYKIN, A. W. (1988). Rhythmic-movement facilitation of learning in working-class Afro-American children. *Journal of Genetic Psychology*, 149, 335–347.
- ALLEN, B. A., & BOYKIN, A. W. (1991). The influence of contextual factors on Afro-American children's performance: Effects of movement opportunity and music. *International Journal of Psychology*, 26, 373–387.
- ANGEL, R., & TIENDA, M. (1982). Headship and household composition among blacks, Hispanics and other whites. *Social Forces*, 61, 508–531.
- AOYAGI, K. (1978). Kinship and friendship in black Los Angeles: A study of migrants from Texas. In D. B. Shimkin, E. M. Shimkin, & D. A. Frate (Eds.), *The extended family in black societies* (pp. 277–355). Chicago: Aldine.
- ASCHENBRENNER, J. (1978). Continuities and variations in black family structure. In D. B. Shimkin, E. M. Shimkin, & D. A. Frate (Eds.), *The extended family in black societies* (pp. 181–200). Chicago: Aldine.
- BARBARIN, O. A. (1983). Coping with ecological transitions by black families: A psychosocial model. *Journal of Community Psychology*, 11, 308–322.
- BARTZ, K. W., & LEVINE, E. S. (1978). Childrearing by black parents: A description and comparison to Anglo and Chicano parents. *Journal of Marriage and the Family*, 40, 709–719.
- BAUMRIND, D. (1971). Current patterns of parental authority. *Developmental Psychology Monographs*, 4(1, Pt. 2).
- BAUMRIND, D. (1972). An exploratory study of socialization effects on black children: Some black-white comparisons. *Child Development*, 43, 261–267.
- BECK, R. W., & BECK, S. H. (1989). The incidence of extended households among middle-aged black and white women: Estimates from a 15-year panel study. *Journal of Family Issues*, 10, 147–168.
- BECK, S. H., & BECK, R. W. (1984). The formation of extended households during middle age. *Journal of Marriage and the Family*, 46, 277–287.
- BECKER, W. (1964). Consequences of different kinds of parental discipline. In M. L. Hoffman & L. W. Hoffman (Eds.), *Review of child development research* (Vol. 1, pp. 169–209). New York: Russell Sage Foundation.
- BELSKY, J., ROBINS, E., & GAMBLE, W. (1984). The determinants of parental competence. In M. Lewis (Ed.), *Beyond the dyad* (pp. 257–280). New York: Plenum.
- BOYKIN, A. W. (1979). Psychological/behavioral verve: Some theoretical explorations and empirical manifestations. In A. W. Boykin, A. J. Franklin, & J. F. Yates (Eds.), *Research directions of black psychologists* (pp. 351–367). New York: Russell Sage Foundation.
- BOYKIN, A. W. (1983). The academic performance of Afro-American children. In J. Spence (Ed.), *Achievement and achievement motives* (pp. 321–371). San Francisco: W. H. Freeman.
- BOYKIN, A. W. (1984). Reading achievement and the social frame of reference of Afro-American children. *Journal of Negro Education*, 53(4), 464–473.
- BOYKIN, A. W., & TOMS, F. D. (1985). Black child socialization. In H. P. McAdoo & J. L. McAdoo (Eds.), *Black children* (pp. 33–52). Beverly Hills, CA: Sage Publications.
- BRADLEY, R., CALDWELL, B., & ROCK, S. (1990). A system for describing elements of the environment. *Early Education and Development*, 1, 237–265.

- BRADLEY, R. H., WHITESIDE, L., CALDWELL, B. M., CASEY, P. H., KELLEHER, K., POPE, S., SWANSON, M., & BARRETT, K. (1993). Maternal IQ, the home environment, and child IQ in low birthweight, premature children. *International Journal of Behavioral Development*, 16(1), 61–74.
- BRODY, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.
- BROMAN, C. L. (1988). Household work and family life satisfaction of blacks. *Journal of Marriage and the Family*, 50, 743–748.
- BRONFENBRENNER, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, 32, 513–531.
- CARTER, E. A., & MCGOLDRICK, M. (1988). Family life cycles and family therapy: An overview. In E. A. Carter & M. McGoldrick (Eds.), *The family life cycle* (2nd ed., pp. 3–28). New York: Gardner Press.
- CAZENAIVE, N. A. (1979). Middle-income black families: An analysis of the provider role. *The Family Coordinator*, 28, 583–593.
- CAZENAIVE, N. A., & SMITH, R. (1990). Gender differences in the perception of black male–female relationships and stereotypes. In H. E. Cheatham & J. B. Stewart (Eds.), *Black families: Interdisciplinary perspectives* (pp. 87–109). New Brunswick, NJ: Transaction.
- COHEN, M. (1984). The ethnomedicine of Garfina (Black Caribs) of Rio Tinto Honduras. *Anthropological Quarterly*, 57, 16–27.
- COHEN, R. (1969). Conceptual styles, cultural conflict, and nonverbal tests of intelligence. *American Anthropologist*, 71, 828–856.
- COLLETTA, N. D. (1979). Support systems after divorce: Incidence and impact. *Journal of Marriage and the Family*, 41, 837–846.
- COLLETTA, N. D. (1981). Social support and risk of maternal rejection by adolescent mothers. *Journal of Psychology*, 109(2), 191–197.
- CUTRIGHT, P. (1971). Income and family events: Family income, family size, and consumption. *Journal of Marriage and the Family*, 33, 161–173.
- DARITY, W., JR., & MYERS, S. L., JR. (1983). Changes in black family structure: Implications for welfare dependency. *American Economic Review Papers and Proceedings*, 73, 59–64.
- DARITY, W., JR., & MYERS, S. L., JR. (1984). Does welfare dependency cause female headship? The case of the black family. *Journal of Marriage and the Family*, 46, 765–779.
- DUNCAN, O. D. (1968). Inheritance of poverty or inheritance of race? In D. P. Moynihan (Ed.), *On understanding poverty* (pp. 116–149). New York: Basic Books.
- DURRETT, M. E., O'BRYANT, S., & PENNEBAKER, J. W. (1975). Childrearing reports of white, black, and Mexican-American families. *Developmental Psychology*, 11, 871.
- DUVALL, E. M. (1971). *Family development*. Philadelphia: Lip-pincott.
- DYKSTRA, R. R., & MANFRA, J. A. (1985). Serial marriage and the origin of the black stepfamily: The Rowonty evidence. *Journal of American History*, 72(1), 18–44.
- ELARDO, R., BRADLEY, R., & CALDWELL, B. (1975). The relation of infant's home environment to mental test performance from 6–36 months. *Child Development*, 46, 71–76.
- EMMERICH, W. (1977). Structure and development of personal-social behaviors in economically disadvantaged preschool children. *Genetic Psychology Monographs*, 95(2), 191–245.
- FLANAGAN, W. G. (1978, August). *The extended family as an agent of social change*. Paper presented at 9th World Congress of the International Sociological Association, Uppsala University, Uppsala, Sweden.
- FOGEL, R., & ENGELMAN, S. (1974). *Time on the cross* (2 vols.). Boston: Little, Brown.
- FRANKLIN, J. H. (1967). *From slavery to freedom: A history of Negro Americans* (3rd ed.). New York: Knopf.
- FURSTENBERG, F., BROOKS-GUNN, J., MORGAN, J. P. (1987). *Adolescent mothers in later life*. New York: Cambridge University Press.
- GARBARINO, J., DUBROW, N., KOSTELNY, K., & PARDO, C. (1992). *Children in danger*. San Francisco: Jossey-Bass.
- GENOVESE, E. D. (1976). *Roll, Jordan, roll*. New York: Random House.
- GLENN, N. D. (1992). What does family mean? *American Demographics*, 14, 30–35.
- GLICK, P., & NORTON, A. J. (1979). Marrying, divorcing, and living together in U.S. today. *Population Bulletin*, 32(5), 1–40.
- GUTMAN, H. G. (1976). *The black family in slavery and freedom, 1750–1925*. New York: Vintage.
- HACKER, A. (1991). *Two nations: Black and white, separate, hostile, and unequal*. New York: Macmillan.
- HALE-BENSON, J. (1986). *Black children: Their roots, culture and learning styles* (rev. ed.). Baltimore, MD: Johns Hopkins University Press.

- HATCHETT, S. J., COCHRAN, D. L., & JACKSON, J. S. (1991). Family life. In J. S. Jackson (Eds.), *Life in black America* (pp. 46–84). Beverly Hills, CA: Sage Publications.
- HERSKOVITS, M. J. (1966). *The myth of the Negro past*. Boston: Beach.
- HILL, R. (1990). Economic forces, structural discrimination, and black family instability. In H. E. Cheatham & J. B. Stewart (Eds.), *Black families: Interdisciplinary perspectives* (87–109). New Brunswick, NJ: Transaction Press.
- HOCHSCHILD, J. L. (1989). Equal opportunity and the estranged poor. *Annals of the American Academy of Political and Social Science*, 501, 143–155.
- HOFFERTH, S. L. (1984). Kin network, race, and family structure. *Journal of Marriage and the Family*, 46, 791–806.
- HURLBERT, J. S. (1990). The effect of marital status on the form and composition of social networks. *Social Science Quarterly*, 71(1), 164–174.
- KELLAM, S. G., ADAMS, R. G., BROWN, C. H., & ENSMINGER, M. E. (1982). The long-term evolution of the family structure of teenage and older mothers. *Journal of Marriage and the Family*, 44, 539–554.
- KELLAM, S. G., ENSMINGER, M. E., & TURNER, R. J. (1977). Family structure and the mental health of children: Concurrent and longitudinal community-wide studies. *Archives of General Psychiatry*, 34, 1012–1022.
- KELLEY, M. L., POWER, T. G., & WIMBUSH, D. D. (1992). Determinants of disciplinary practices in low-income black mothers. *Child Development*, 63, 573–582.
- KELLEY, M. L., SANCHEZ-HUCLES, J., & WALKER, R. R. (1993). Correlates of disciplinary practice in working- to middle-class African-American mothers. *Merrill-Palmer Quarterly*, 39(2), 252–264.
- LANDRY, B., & JENDREK, M. S. (1978). The employment of wives in middle-class black families. *Journal of Marriage and the Family*, 42, 787–797.
- LEE, G. R. (1977). *Family structure and interaction*. New York: Lippincott.
- LESLIE, L. A., & ANDERSON, E. A. (1988). Men's and women's participation in domestic roles. *Journal of Family Psychology*, 2(2), 212–226.
- LIEBERSON, S. (1980). *A piece of the pie*. Berkeley: University of California Press.
- LIEBERSON, S., & CARTER, D. K. (1979). Making it in America: Differences between black and white ethnic groups. *American Sociological Review*, 44, 347–366.
- MACCOBY, E. E. (1980). *Social development: Psychological growth and the parent-child relationship*. New York: Harcourt Brace Jovanovich.
- MARTIN, E. P., & MARTIN, D. (1978). *The black extended family*. Chicago: University of Chicago Press.
- MCADOO, H. P. (1978). Factors related to stability in upwardly mobile black families. *Journal of Marriage and the Family*, 40, 761–776.
- MCADOO, H. P. (1991). Family values and outcomes for children. *Journal of Negro Education*, 60, 361–365.
- MCADOO, H. P. (1992). A review of the literature related to family therapy in the black community. *Journal of Contemporary Psychotherapy*, 9(1), 15–19.
- MCADOO, J. L. (1988). Changing perspectives on the role of black fathers. In P. Bornstein & C. P. Cowan (Eds.), *Fatherhood today: Men's changing role in the family* (pp. 79–93). New York: Wiley.
- MCADOO, J. L. (1993). The role of African-American fathers: An ecological perspective. *Families in Society: The Journal of Contemporary Human Services*, 74(1), 28–35.
- MCLANAHAN, S. S., ASTONE, N. M., & MARKS, N. F. (1991). The role of mother-only families in reproducing poverty. In A. C. Huston (Ed.), *Children in poverty* (pp. 51–78). New York: Cambridge University Press.
- MCLANAHAN, S., & GARFINKEL, I. (1989). Single mothers, the underclass, and social policy. *Annals of the American Academy of Political and Social Science*, 501, 92–105.
- MCLANAHAN, S. S., WEDEMEYER, N. V., & ADELBERG, T. (1981). Network structure, social support, and psychological well-being in single-parent family. *Journal of Marriage and the Family*, 43, 601–612.
- MEACHAM, M. (1983). The myth of the black matriarchy under slavery. *Mid-American Review of Sociology*, 8, 23–41.
- MOORE, E. G. J. (1985). Ethnicity as a variable in child development. In M. B. Spencer, G. K. Brookins, & W. R. Allen (Eds.), *Beginnings: The social and affective development of black children* (pp. 101–115). Hillsdale, NJ: Erlbaum.
- MOORE, E. G. J. (1986). Family socialization and the IQ-test performance of traditionally and transracially adopted black children. *Developmental Psychology*, 22(3), 317–326.
- MOORE, E. G. J. (1987). Ethnic social milieu and black children's intelligence test achievement. *Journal of Negro Education*, 56(1), 44–52.
- MYERS, H. F. (1982). Research on the Afro-American family: A critical review. In B. A. Bass, G. E. Wyatt, & G. J. Powell (Eds.), *The Afro-American family* (pp. 35–68). New York: Grune and Stratton.

- NATHAN, R. P. (1989). Institutional change and the challenge of the underclass. *Annals of the American Academy of Political and Social Science*, 501, 170–181.
- OGBU, J. U. (1988). Cultural diversity and human development in black children and poverty: A developmental perspective. In D. T. Slaughter (Ed.), *New directions for child development*, 42. San Francisco: Jossey-Bass.
- OLIVER, M. L. (1988). The urban black community as network: Toward a social network perspective. *Sociological Quarterly*, 29, 623–645.
- OSBORNE, R. T., & MCGURK, F. C. J. (EDS.). (1982). *Testing of Negro intelligence*. Athens, GA: The Foundation for Human Understanding.
- OTTO, J. J., & BURNS, A. M., III. (1983). Black folks and poor buckras: Archeological evidence of slave and overseer living conditions on an antebellum plantation. *Journal of Black Studies*, 14, 185–200.
- PEARSON, J. L., HUNTER, A. G., ENSMINGER, M. E., & KELLAM, S. G. (1990). Black grandmothers in multigenerational households: Diversity in family structure and parenting involvement in the Woodlawn community. *Child Development*, 61(2), 434–442.
- PERLMANN, J. (1988). *Ethnic differences*. New York: Cambridge University Press.
- RAMEY, C. T., & CAMPBELL, F. (1987). The Carolina Abecedarian project: An educational experiment concerning human malleability. In J. J. Gallagher & C. T. Ramey (Eds.), *The malleability of children* (pp. 115–127). Baltimore, MD: Paul H. Brookes.
- RAMEY, C. T., & RAMEY, S. L. (1992). Effective early intervention. *Mental Retardation*, 30(6), 337–345.
- REID, J. (1982). Black America in the 1980's. *Population Bulletin*, 37(4), 1–37.
- RUTLEDGE, E. (1988). African-American socialization experiences by family structure. *Journal of Black Studies*, 19(2), 204–215.
- SAMEROFF, A. J., SEIFER, R., BALDWIN, A., & BALDWIN, C. (1993). Stability of intelligence from preschool to adolescence: The influence of social and family risk factors. *Child Development*, 64(1), 80–97.
- SAMUDA, S. J. (1974). *Psychological testing of American minorities*. New York: Harper & Row.
- SCANZONI, L. D., & SCANZONI, J. (1981). *Men, women and change: A sociology of marriage*. New York: McGraw-Hill.
- SCARR, S., & MCCARTNEY, K. (1983). How people make their own environments. *Child Development*, 54, 424–435.
- SLAUGHTER, D. T., & DILWORTH-ANDERSON, P. (1988). Care of black children with sickle cell disease: Fathers, maternal support, and esteem. *Family Relations*, 37, 281–287.
- SLESINGER, D. P. (1980). Rapid changes in household composition among low-income mothers. *Family Relations*, 29, 221–228.
- STACK, C. (1974). *All our kin: Strategies for survival in the black community*. New York: Harper & Row.
- STERNBERG, R. (1988). A triarchic view of intelligence in cross-cultural perspective. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context* (pp. 60–87). New York: Cambridge University Press.
- TESTA, M., ASTONE, M., KROGH, M., & NECKERMAN, K. M. (1989). Employment and marriage among inner-city fathers. *Annals of the American Academy of Political and Social Science*, 501, 79–91.
- TIENDA, M., & ANGEL, R. (1982). Determinants of extended household structure: Cultural pattern or economic need? *American Journal of Sociology*, 87, 1360–1383.
- TOLSON, T. F. J., & WILSON, M. N. (1990). The impact of two- and three-generational black family structure on perceived family climate. *Child Development*, 61, 416–428.
- U.S. BUREAU OF THE CENSUS. (1993). *Household and family characteristics: March 1992* (Current Population Report Series P-20, No. 447). Washington, DC: U.S. Government Printing Office.
- WACQUANT, L. J. D., & WILSON, W. J. (1989). Cost of racial and class exclusion in the inner city. *Annals of the American Academy of Political and Social Science*, 501, 8–25.
- WICKER, A. W. (1979). *An introduction to ecological psychology*. Monterey, CA: Brooks/Cole.
- WILSON, M. N. (1986). The black extended family: An analytical consideration. *Developmental Psychology*, 22, 246–258.
- WILSON, M. N. (1989). Child development in the context of the black extended family. *American Psychologist*, 44, 380–385.
- WILSON, M. N., HINTON, I. D., TOLSON, T. F. J., SIMMONS, F., STAPLES, W., ASKEW, T., & MCKIM, L. (1990). *An analysis of adult-child interaction patterns in diverse African-American families* (Report No. PS-018-964). Charlottesville, VA: University of Virginia (ERIC Document Reproduction Service No. ED, 321, 893).
- WILSON, M. N., HINTON, I. D., BOYER, J., PHILIP, D., MCELROY, A., SMITH, U., & SOLOMON, U. (1992). The effect of family composition on parent-child interaction. *International Journal of Psychology*, 27(3–4), 252.
- WILSON, M. N., & TOLSON, T. F. J. (1986). A social interac-

- tion analysis of two- and three-generational black families. In P. Dail & R. Jewson (Eds.), *In praise of fifty years: Groves Conference on the Conservation of Marriage and the Family* (pp. 43–53). Lake Mills, IA: Graphic Publishing.
- WILSON, W. J. (1991). Public policy research and the truly disadvantaged. In C. Jencks & P. E. Peterson (Eds.), *The urban underclass* (pp. 460–482). Washington, DC: Brookings Institution.
- WOJTKIEWICZ, R. A., MCLANAHAN, S. S., & GARFINKEL, I. (1990). The growth of families headed by women: 1950–1980. *Demography*, 27(1), 14–30.
- YOUNG, V. H. (1970). Family and childhood in a southern Negro community. *American Anthropologist*, 72, 269–288.
- YOUNG, V. H. (1974). A black American socialization pattern. *American Ethnologist*, 1, 405–412.
- ZOLLAR, A. C. (1985). *A member of the family*. Chicago: Nelson-Hall.
- ZOLLAR, A. C., & HONNOLD, J. (1988). Socioeconomic characteristics and kin interaction in black middletown. *Western Journal of Black Studies*, 12, 9–18.

MELVIN N. WILSON

AGE AND THE CONTENT OF INTELLIGENCE TESTS

In everyday life, the term *intelligence* is often used to aid in understanding one's own or another's behavior or underlying competence. Ideas about intelligence and whether or not intelligence changes with age not only affect estimates of skills at present but also influence decisions people make about goals and future plans.

Indeed, there are many perspectives on defining intelligence, each of which influences the specific content of tests said to measure this construct. J. M. Sattler (1992) discussed more than twenty separate, yet somewhat overlapping definitions of intelligence, many of which have led to the development of distinct scales to measure a variety of skills and abilities. Because definitions of intelligence vary, and because intelligence is an ability or cluster of abilities whose existence must be inferred on the basis of test performance, one must be extremely cautious about discussing what intelligence is and how it changes with age, based on a specific test's content.

Most definitions of intelligence used to determine

test content emphasize (1) the manipulation of figures or numbers; (2) the ability to think abstractly; (3) the ability to form relationships between words, objects, or numbers; (4) skill in adapting to novel environments; (5) the ability to apply one's past experience to the solution of a problem; and (6) the ability to adapt to real-world environments that are relevant to everyday life (Sattler, 1992, p. 45). Additionally, most test constructors have some working definition of intelligence in mind, and though they might disagree over the extent to which intelligence is a general ability, most nevertheless assume that it is to a certain extent multidimensional.

Psychometric approaches to intelligence rely on "factor analysis," a statistical technique that provides an empirical guide to how many underlying factors explain relationships among a number of scales measuring various abilities and skills. Depending on what specific scales are used and the nature of one's sample (whether it is a very diverse one or not), several common factors or just one could be derived from factor analysis. The "factors" identified by factor analysis are merely statistical abstractions, as when one labels factors as "verbal" or "performance" intelligence.

TEST CONTENT AND DEVELOPMENTAL CHANGE

Measures of intelligence are quite numerous and almost as variable as its definition. For the present purpose, the issue of test content and its change with age will be discussed with reference to the major scales in use today: the Wechsler scales and the Stanford-Binet scale.

Factor-analytic studies of the Wechsler scales for both children and adults and the latest edition of the Stanford-Binet suggest that each scale's content is best understood in terms of multiple factors, based on the standardization samples of children and young adults aged 2–23. Stability of factor structure across age seems to be the rule (Sattler, 1992).

For the revised Wechsler Preschool and Primary Scale of Intelligence (WPPSI-R) (Wechsler, 1989), FACTOR ANALYSIS of the data from the standardization sample aged 3–7 yields a verbal factor and a performance factor; for all WPPSI-R subtests, loadings on *g* are relatively high (Sattler, 1992).

For the Wechsler Intelligence Scale for Children, or WISC-III (Wechsler, 1991), these factors are verbal comprehension, perceptual organization, and processing speed (Sattler, 1992). "Verbal comprehension" measures verbal knowledge and understanding obtained informally or through formal education, and reflects the use of verbal skills in new situations. "Perceptual organization" measures the ability to interpret and organize visually presented material within a time limit. "Processing speed" measures the ability to maintain attention or concentrate in processing information rapidly while scanning an array. While verbal comprehension is indexed by most of the verbal subtests (vocabulary, information, similarities, comprehension), it is also measured to a certain extent by "picture completion" (identifying essential missing parts of pictures) and "picture arrangement" (arranging picture into meaningful sequences). The performance subtests of the WISC-III (Block Design, or reproducing stimulus designs with multicolored blocks; object assembly, or puzzles; and picture completion) primarily index "perceptual organization," as do "mazes" and "picture arrangement," to a certain extent. "Arithmetic," "digit span" (repeating numbers in either forward or backward order that one has just heard), "coding" (copying symbols from a key), and "symbol search" (deciding whether a symbol appears in an array) reflect processing speed. Performance on each subtest of the WISC-III can also be explained to a certain extent by a general intelligence factor (Sattler, 1992).

Over a twelve-year (age 5½ to 16½) period, correlations both within and between the verbal and performance scales derived from the earlier versions of the WPPSI and WISC are extremely high, ranging from .73 to .92 across studies (Bishop & Butterworth, 1979; Tew & Lawrence, 1983; Yule, Gold, & Busch, 1982). Thus, based on the content of the earlier versions of the current Wechsler scales, there is a high degree of stability in intelligence over many years in childhood and adolescence.

The factor structure of the Wechsler Adult Intelligence Scale, or WAIS-R (Wechsler, 1981), is similar to that of the second edition of the WISC (WISC-R): verbal comprehension, perceptual organization, and freedom from distractibility, except that at age 18-19, picture completion fails to load on perceptual organi-

zation. For the WAIS-R, the verbal subtests are somewhat better measures of *g* than are the performance subtests. Numerous cross-sectional analyses clearly suggest that verbal subtests exhibit less age-related decline than do the performance subscales (Botwinick, 1984). The extent to which age cohorts differ in terms of levels of education principally accounts for the age effects in verbal skills relative to performance skills (Kaufman, 1990). Analyses of standardization data for the Stanford-Binet yield a similar picture of differential age (cohort) decline (Kausler, 1991).

For the fourth edition of the Stanford-Binet, or SB IV, factors derived from its content based on data from the standardization sample ranging in age from 2 to 23 are (1) from age 2 to 6, verbal comprehension and nonverbal reasoning/visualization, while (2) from ages 7 to 23, these factors are similar, except for a third factor termed "memory." While verbal comprehension is defined as above in the Wechsler scales, "nonverbal reasoning/visualization" reflects the ability to interpret and organize visually perceived material, arithmetic skill with the use of verbal or visual cues, the use of reasoning to solve problems, pattern visualization, and visual-motor skills. Memory reflects attention or the ability to concentrate, as well as the ability to produce and understand sequences (Sattler, 1992). As above, all SB IV subtests load in varying degrees on a general-intelligence factor. While long-term stability of the factor scores has not been established, reliabilities are high in each case (Sattler, 1992).

APPROACHES TO INTELLIGENCE AND CHANGE WITH AGE

While the content of most intelligence tests is not closely linked to a particular view about how intelligence is defined and organized, at this more general level as well, change with age can nevertheless be discussed. Ideas about how intelligence is organized vary greatly. W. Stern's (1914) unifactor theory suggests that individuals differ in terms of the degree to which they possess GENERAL INTELLIGENCE. Charles SPEARMAN (1904) suggested that intelligence was best explained by two factors, *g*, or general intelligence, and *s*, or test-specific abilities. Tests with high *g*-to-*s* ratios reflect the "eduction of relations and correlates." Subse-

quently, Cyril BURT, a British psychologist, proposed a hierarchical model of intelligence that specified four factors, differing in terms of their generality (Sattler, 1992). The general-intelligence factor is broadest, in addition to two group factors (termed verbal-educational and spatial-mechanical) that are more specific, followed by specific (task) factors, and, finally, chance factors.

A three-level hierarchical model of intelligence has also been proposed by J. E. Gustafsson (1984). It incorporates not only general ability but also the secondary factors of FLUID AND CRYSTALLIZED INTELLIGENCE (see below), as well as general visualization followed by primary abilities.

While theories of intelligence typically have not translated into specific test content, a three-factor hierarchical model has guided the development of SB IV (Thorndike, Hagen, & Sattler, 1986). This model has *g* at the highest level, followed by crystallized, fluid, and memory factors at the second level, with more specific factors (e.g., verbal, quantitative, and abstract verbal reasoning) at the lowest level.

The most complex of approaches to intelligence is that of J. P. GUILFORD (1967), termed the STRUCTURE-OF-INTELLECT MODEL (SI) of intelligence, which differentiates *operations*, the basic psychological processes involved in ability formation (e.g., memory, cognition, awareness, knowing, convergent and divergent production); (2) *contents*, the nature of the material dealt with (number, letters or words, behaviors); and (3) *products*, the consequence of the interaction of operations acting on content. As the above three classes of abilities are seen as independent, the resulting model theoretically allows for 150 factors. While Guilford (1985) reorganized the theory to specify a hierarchical ability model, research does not support the existence of all 150 abilities, nor is the hypothesized independence among such factors supported (Brody, 1992).

While the SI model has, for the most part, yet to be integrated into developmental research in intelligence, the major exception are studies of “divergent thinking,” where cross-sectional comparisons of younger and older adults suggest a decline in divergent thinking across age levels (see Albaugh et al., 1982).

Despite the popularity of factorially complex ideas about intelligence, many scholars nevertheless adhere

to a generalist position (Brody, 1992; Rebok, 1987; Scarr, 1989). By contrast, the adult-development literature clearly suggests that intelligence is multidimensional (Carroll, 1993; Kausler, 1991; Schaie, 1990).

CRYSTALLIZED AND FLUID ABILITIES

The distinction between crystallized (G_c) and fluid (G_f) abilities is especially suited to development in that both intelligences are defined in such a way that predictions about developmental change are possible. Both G_f and G_c help to define one's intelligence quotient (IQ), precluding generalized comparisons across age.

Fluid and crystallized abilities have been formally defined as the “process of perceiving relationships, educating correlates, maintaining span of immediate awareness in reasoning, abstracting, concept formation, and problem solving” (Horn, 1978, p. 220). Unspeeded and speeded tasks involving figures, symbols, or words can measure G_f , and performance is relatively independent of intensive or extended education and acculturation (Horn, 1978). By contrast G_c reflects “relatively advanced education and acculturation either in the fundamentals (contents) of the problem or in the operations that must be performed on the fundamentals” (Horn, 1978, pp. 221–222). Crystallized skills come about (crystallize) as a function of more organized, more systematic, acculturated learning, but G_f is fluid, or fluctuates with the demands made on one in novel situations. Thus, where a problem demands a novel response, G_f will come into play, whereas when previously learned skills are required, G_c will be called on. Horn (1978) suggested that this distinction is not always clear-cut, as when the task could require the exercise of either general ability.

Decreased neurophysiological functioning with age is said to influence G_f , while cumulative intensive acculturation/education influences G_c . The evidence for the former is indirect and somewhat sketchy (Brody, 1992; Horn, 1982, 1985). Other influences involve selective learning, family size and composition, values of parents or peers, one's attitude toward aging, and being labeled by others as “disadvantaged” or “old.” It is thought that G_f increases up until late adolescence and then declines thereafter, whereas g_c should generally increase or remain stable over the adult years

(Cattell, 1963; Horn, 1978; Horn & Cattell, 1966, 1967).

John Horn (1982, 1985) related G_f and G_c to measures of personality (carefulness), sensory/perceptual-motor slowing, short-term memory, and attention, to explain more fully those processes that contribute to intellectual functioning. This revised approach is hierarchical, from sensory functioning, the most specific, to thinking, the most general.

To a large extent, the above factor-analytic findings for the Wechsler scales and the Stanford-Binet can be understood in terms of the distinction between fluid (encompassing the factors of nonverbal reasoning, perceptual organization, and freedom from distractibility) and crystallized (verbal comprehension) ability.

PRIMARY MENTAL ABILITIES

L. L. THURSTONE'S (1938) theory of primary mental abilities (PMA) specifies several factors: spatial ability, perceptual speed, numerical ability, verbal relations, words, memory, and induction. While the PMA approach has not been historically important in the assessment of child and adolescent intelligence, K. Warner Schaie (1979, 1990) used the PMA theory to study adult intellectual development.

Schaie's twenty-eight-year longitudinal data suggest that intellectual abilities increase through one's forties, relative stability for most skills exists through one's fifties and early sixties, and average losses become significant for most abilities thereafter. Such declines are greater for persons with cardiovascular illness, where performance is speeded; for those who are poorly educated; and for those who live in intellectually depriving environments. Moreover, there are individual differences in the extent of decline; global loss is rare. Less than one-third experience decrement until age 74, and only 30–40 percent experience significant losses in intellectual skills by age 81 (Schaie, 1990).

One of the most important findings in adult intelligence is the influence of cohort effects on intellectual functioning, independently of age (maturation) (Kaufman, 1990; Schaie, 1990). Schaie's data clearly suggest that cohort effects are at least as important as maturation as an influence on intelligence in adulthood, in that they affect the baseline reference point from which age-related changes in intelligence can be

understood. For some abilities, cohort differences are positive (younger cohorts perform more adequately), as in the case of the PMA verbal meaning, spatial orientation, and inductive reasoning (Schaie, 1990), largely because of higher levels of education and better health for such persons. For other skills, cohort effects are negative (younger cohorts perform less adequately), as in the case of PMA number skill or word fluency.

Schaie's findings are in contrast to a more generalist orientation, as measured by the Army Alpha test (Jones & Conrad, 1933) or the WAIS-R. While Army Alpha scores demonstrate negative age effects, as noted above, verbal subscales of the WAIS-R exhibit comparatively little age decrement relative to the performance subscales (Kaufman, 1990; Kausler, 1991).

THE INFORMATION-PROCESSING APPROACH

The information-processing approach to intelligence (Sternberg, 1985; Sternberg & Detterman, 1979) envisions an even more tenuous direct relationship to existing scale content. It stresses the person as an active processor of information contained in a problem or in the real world. Individuals develop essential logical operations and strategies by which to understand and analyze information presented to them. In this case, the speed or accuracy of processing simple stimuli are of interest. Consequently, basic abilities such as reaction time, inspection time necessary to identify a simple visual or auditory stimulus, attentional processes, speed of information processing, and stimulus discrimination have been studied.

Research in this area focuses on specific tasks or types of items commonly found in most intelligence tests, such as spatial relations, analogies, and block design. This approach focuses on essential component processes that are themselves a function of the interaction between task influences and person influences, such as encoding, storage, retrieval, rule formation, and pattern analysis. Typically, the identification of such processes is accomplished through a task or component analysis, whereby performance on an intelligence-test item is broken down into more-basic units (processes) that can be studied thoroughly (Brody, 1992).

J. C. Campione and A. L. Brown (1978) and F. G. Borkowski (1985) offered an information-processing perspective that has two basic components, the architectural system and the executive system. The former refers to biologically based skills that are necessary for processing information, such as immediate memory span, retention of stimulus traces, and efficiency of information processing, and the latter refers to acquired skills that guide problem solving, including one's knowledge base, schema, control processes, and metacognition. The architectural system parallels fluid abilities, and the executive system seems to represent crystallized skills.

Deficits in most component processes underlying fluid ability have been found with increasing age (Brody, 1992; Kausler, 1991). Considerable debate surrounds the extent to which multiple basic cognitive processes reflect general intelligence.

JOB ANALYSIS AND INTELLIGENCE

A somewhat unique approach to intelligence is that based on job analysis. By ascertaining the tasks that must be performed to complete a job successfully, criteria can be developed so that prospective employees can be selected on the basis of their ability to meet these criteria. Job analyses may be specifically job-oriented, wherein tasks necessary to do the job are identified, or they can be worker-oriented, wherein jobs are expressed in terms of abilities required to do work, such as visual or auditory ability, or information processing (Muchinsky, 1990). Where inferences about intelligence are being made, they may reflect the utilization of previously acquired skills (*Gc*) or the capacity to solve novel problems (*Gf*). In addition, specific domains of intelligence as they relate to work may be specified, such as verbal, numerical, mechanical, or spatial ability (Miner, 1992). The influence of verbal ability pervades occupational levels, but such is not the case for numerical, mechanical, or spatial skills, whose relevance is job-specific (Miner, 1992).

PIAGETIAN INTELLIGENCE

Jean PIAGET (1963) and Piaget and B. Inhelder (1969) have suggested that intellectual development through adolescence progresses through a series of discrete biologically based stages: sensorimotor (birth–

2 years), preoperational (2–7 years), concrete operations (7–11 years), and formal operations (11 and beyond). As essential mental processes are transformed via the interaction of the organism with the environment, they evolve from those that are sensory-perceptual and survival-oriented to those that enable the individual to communicate with others as well as represent, manipulate, and understand the environment. What distinguishes later childhood and early adolescence from infancy and early childhood is “operational thought,” characterized by the ability to use symbols (words) to solve problems and perform various mental activities. Concrete operational children can logically solve problems with which they have had experience, but formal-operational thinking, characteristic of adolescents, is more abstract. Such persons can logically reason and solve hypothetical problems.

Although no standardized Piagetian battery exists, this approach has influenced the development of sensorimotor scales with which to assess infant intelligence (Sattler, 1992; Uzgiris & Hunt, 1975), in contrast to scales designed for older children and adults, whose content is more verbal/abstract in nature. Additionally, correlations between such Piagetian tasks as the pendulum problem (assessing formal operations) or conservation tasks are in the moderately positive range (Sattler, 1992).

Under the assumption that older people regress to an earlier level of development, many have investigated Piagetian intelligence among older adults. Whether such regression occurs is difficult to ascertain because nearly all studies are cross-sectional in nature (Papalia & Bielby, 1974). Nevertheless, these studies clearly reveal that older individuals are less able to solve tasks requiring the transformation of number of objects or to deal with changes in the weight or volume of different objects. Deficits in Piagetian task performance are rarely found in healthy, educated elderly persons, and errors on these tasks are also common among younger persons (Blackburn & Papalia, 1992; Kausler, 1991).

J. M. Rybash, W. J. Hoyer, and P. A. Roodin (1986) suggest that formal-operational thinkers ignore the context in which the problem to be solved is embedded. Consequently, abstract thinking is overemphasized, and the importance of emotion in making everyday decisions is deemphasized (Blanchard-Fields,

1986). S. J. Ceci, N. N. Nightingale, and J. G. Baker (1992) also emphasize the importance of contextual influences on the intellectual functioning of children, wherein the sociocultural features of the situation, such as the age or sex-role appropriateness of the task, as well as the presence or absence of peers or familiar persons, also influence performance.

Some theories, such as those of Piaget, Thurstone, or Horn and Cattell, seem to parallel current test content and have led to the investigation of developmental change, but others, such as Guilford's, have not. In this respect, examining intelligence in terms of basic cognitive processes that may underlie more general abilities in the context of genetic and acquired motivational or interpersonal influences may be most fruitful and consequently lead to an integration of a variety of seemingly incongruent approaches within a developmental framework.

Although the bulk of evidence suggests that intelligence is multidimensional, researchers are hardly in agreement regarding what dimensions are both necessary and sufficient for clear operational definitions that might lead to the development of specific test content. Yet, parallels may exist between fluid and crystallized abilities and the process and content dimensions of Guilford. Likewise, both the attentional and essential component processes inherent in fluid ability, as well as in Piagetian operational and postformal thought, may be fruitful areas for integrative research. In addition, conducting task analyses and specifying the contextual influences on fluid processes are likely candidates for intervention research with individuals of all ages. As individuals are both active and passive with respect to their environment, the seemingly diverse ideas about intelligence and its development can easily coexist.

(See also: AGING AND INTELLIGENCE.)

BIBLIOGRAPHY

- ALBAUGH, P., PARHAM, I., COLE, K., & BERRÈN, J. E. (1982). Creativity in adulthood and old age: An exploratory study. *Educational Gerontology*, 8, 101-116.
- BISHOP, D., & BUTTERWORTH, G. E. (1979). A longitudinal study using the WPPSI and WISC-R with an English sample. *British Journal of Educational Psychology*, 49, 156-168.
- BLACKBURN, J., & PAPALIA, D. (1992). *The study of adult cognition from a Piagetian perspective*. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 141-160). New York: Cambridge University Press.
- BLANCHARD-FIELDS, F. (1986). Reasoning in adolescence and adults on social dilemmas varying in emotional saliency. *Psychology and Aging*, 1, 325-333.
- BORKOWSKI, F. G. (1985). Signs of intelligence: Strategy generalization and metacognition. In S. R. Yussen (Ed.), *The growth of reflection in children* (pp. 105-144). Orlando, FL: Academic Press.
- BORNSTEIN, M. H. (1989). Stability in early mental development: From attention and information processing in infancy to language and cognition in childhood. In M. Bornstein & N. Krasnegor (Eds.), *Stability and continuity in mental development* (pp. 140-172). Hillsdale, NJ: Erlbaum.
- BOTWINICK, J. (1984). *Ageing and behavior*. New York: Springer-Verlag.
- BRODY, N. (1992). *Intelligence*. New York: Academic Press.
- CAMPIONE, J. C., & BROWN, A. L. (1978). Toward a theory of intelligence: Contributions from research with retarded children. *Intelligence*, 2, 279-304.
- CARROLL, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. Cambridge: Cambridge University Press.
- CATTELL, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1-22.
- CECI, S. J., NIGHTINGALE, N. N., & BAKER, J. G. (1992). The ecologies of intelligence: Challenges to traditional views. In D. K. Detterman (Ed.), *Current topics in human intelligence* (Vol. 2, pp. 61-82). Norwood, NJ: Ablex.
- DETERMAN, D. (1987). What does reaction time tell us about intelligence? In P. A. Vernon (Ed.), *Speed of information processing and intelligence* (pp. 10-31). Norwood, NJ: Ablex.
- GUILFORD, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- GUILFORD, J. P. (1985). The structure of intellect model. In B. B. Wolman (Ed.), *Handbook of intelligence* (pp. 225-266). New York: Wiley.

- GUSTAFSSON, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179–203.
- HORN, J. L. (1978). Human ability systems. In P. B. Baltes (Ed.), *Life-span development and behavior* (Vol. 1, pp. 211–256). New York: Academic Press.
- HORN, J. L. (1982). The aging of human abilities. In B. B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 847–870). Englewood Cliffs, NJ: Prentice-Hall.
- HORN, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence* (pp. 267–300). New York: Wiley.
- HORN, J. L., & CATTELL, R. B. (1966). Refinement and test of the theory of fluid and crystallized general intelligence. *Journal of Educational Psychology*, 53, 253–270.
- HORN, J. L., & CATTELL, R. B. (1967). Age differences in fluid and crystallized intelligence. *Acta Psychologica*, 26, 107–129.
- JENSEN, A. (1987). Individual differences in the Hick paradigm. In P. A. Vernon (Ed.), *Speed of information processing and intelligence* (pp. 32–50). Norwood, NJ: Ablex.
- JONES, H. E., & CONRAD, H. S. (1933). The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty. *Genetic Psychology Monographs*, 13, 223–298.
- KAUFMAN, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn & Bacon.
- KAUSLER, K. H. (1991). *Experimental psychology, cognition, and human aging*. New York: Springer-Verlag.
- MINER, J. B. (1992). *Industrial organizational psychology*. New York: McGraw-Hill.
- MUCHINSKY, P. M. (1990). *Psychology applied to work*. Pacific Grove, CA: Brooks/Cole.
- PAPALIA, D., & BIELBY, D. (1974). Cognitive functioning in middle-aged and elderly adults: A review of research based on Piaget's theory. *Human Development*, 17, 424–433.
- PIAGET, J. (1963). *The psychology of intelligence*. New York: Littlefield-Adams.
- PIAGET, J., & INHELDER, B. (1969). *The psychology of the child*. New York: Basic Books.
- REBOK, G. (1987). *Life span cognitive development*. New York: Holt, Rinehart & Winston.
- RYBASH, J. M., HOYER, W. J., & ROODIN, P. A. (1986). *Adult cognition and aging*. Elmsford, NY: Pergamon Press.
- SATTLER, J. M. (1992). *Assessment of children* (3rd ed.—rev.). San Diego, CA: Jerome M. Sattler.
- SCARR, S. (1989). Protecting general intelligence: Constructs and consequences for interventions. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 74–118). Urbana: University of Illinois Press.
- SCHAIK, K. W. (1979). The primary mental abilities in adulthood: An exploration in the development of psychometric intelligence. In P. B. Baltes & O. Brim (Eds.), *Life-span development and behavior* (Vol. 2, pp. 68–115). New York: Academic Press.
- SCHAIK, K. W. (1990). Adult intellectual development. In J. E. Berrèn & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (pp. 292–309). New York: Academic Press.
- SPEARMAN, C. (1904). General intelligence objectively defined and measured. *American Journal of Psychology*, 15, 72–101.
- STANKOV, L. (1988). Aging, attention, and intelligence. *Psychology and Aging*, 3, 59–74.
- STERN, W. (1914). The psychological methods of testing intelligence. *Educational Psychological Monographs*, 3.
- STERNBERG, R. J. (1985). Cognitive approaches to intelligence. In B. B. Wolman (Ed.), *Handbook of human intelligence* (pp. 59–118). New York: Wiley.
- STERNBERG, R. J., & DETTERMAN, D. (1979). *Human intelligence: Perspectives on its theory and measurement*. Norwood, NJ: Ablex.
- TEW, B. J., & LAWRENCE, K. M. (1983). The relationship between spina bifida children's intelligence scores on school entry and at school leaving: A preliminary report. *Child: Care, Health, and Development*, 9, 13–17.
- THORNDIKE, R. L., HAGEN, E. P., & SATTLER, J. M. (1986). *Technical manual: Stanford-Binet Intelligence Scale* (4th ed.). Chicago: Riverside.
- THURSTONE, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- UZGIRIS, I. C., & HUNT, J. M. (1975). *Assessment in infancy: Ordinal scales of psychological development*. Urbana: University of Illinois Press.
- WECHSLER, D. (1981). *WAIS-R Administration and Scoring Manual*. New York: Psychological Corporation.
- WECHSLER, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence* (rev. ed.). San Antonio, TX: The Psychological Corporation.
- WECHSLER, D. (1991). *Manual for the Wechsler Intelligence Scale for Children* (3rd ed.). San Antonio, TX: The Psychological Corporation.
- YULE, W., GOLD, R. D., & BUSH, C. (1982). Long-term pre-

dictive validity of the WPPSI: An 11-year follow-up study. *Personality and Individual Differences*, 3, 65–71.

BERT HAYSLIP, JR.

AGING AND INTELLIGENCE Intellectual aging is not a unitary process; both individuals and abilities differ in the way they age. Shining examples of older persons who exhibit high intellectual creativity, such as Goethe or Sophocles, stand in contrast to individuals whose intellectual capacities are greatly diminished in old age. A similar contrast exists between different intellectual abilities. For example, if one looks at logical problem solving, one is likely to find decline with advancing age. However, if one looks at word knowledge or aspects of social intelligence, stability or even select advances into old age are possible. There are many faces to the aging of human intelligence.

This diversity in the aging of intelligence and in the aging of individuals is reflected by the existence of distinct and not always compatible research perspectives and interpretative frameworks. For the period of adulthood and old age, three perspectives are of major importance: the *psychometric* (Horn & Hofer, 1992), *information-processing* (Craik, 1983; Salthouse, 1991b), and *expertise* (Charness, 1989; Ericsson & Smith, 1991). In addition, efforts to reconceptualize the meaning and measurement of intelligence continue to capture more appropriately the nature of intellectual activity displayed by aging adults (Alexander & Langer, 1990; Perlmutter, 1990). The search for postformal operations (Labouvie-Vief, 1992), the investigation of practical intelligence (Sternberg & Wagner, 1986), and the measurement of wisdom (Baltes, Smith, & Staudinger, 1992) are three examples of this line of scholarship.

The different research traditions and their associated bodies of knowledge need to be considered jointly to obtain a reasonably full view of the aging mind, a view that reflects the gains and losses of adult intellectual development as well as interindividual and historical variability (Baltes, 1993). The psychometric evidence offers a multidimensional, quantitative picture of age gradients in intellectual abilities and serves to identify antecedents and correlates of individual differences in intellectual functioning among older adults. The information-processing approach aims at a more

precise analysis of aging-associated changes in components of cognitive functioning and their operative interactions. The expertise approach illuminates the ways in which knowledge and practice enrich the mind and helps to identify and analyze instances of intellectual growth and high-level functioning (peak performance). In combination, the expertise and information-processing approaches serve to understand better the interplay between processing limitations and knowledge increments with advancing age.

ADULT AGE GRADIENTS IN PSYCHOMETRIC INTELLIGENCE

Following early work on the life-span development of intelligence by Francis GALTON, Quetelet, and Tettens (cf. Dixon & Baltes, 1986) and based on the development of psychometric tests of intelligence around the turn of the twentieth century, a series of first cross-sectional studies on the aging of intelligence appeared in the 1920s and 1930s. For example, Jones and Conrad (1933) administered the Army Alpha test battery (see ARMY ALPHA AND BETA TESTS OF INTELLIGENCE) to more than 600 individuals in the range of 19 to 60 years of age. Negative age differences were pronounced on tests such as following directions, common sense, number series, and verbal analogies tests, but were small or nonsignificant on arithmetic, antonym-synonym, disarranged sentences, and general information. The existence of differential age gradients for different intellectual abilities was further confirmed with the development of the Wechsler Adult Intelligence Scale (WAIS; see WECHSLER SCALES OF INTELLIGENCE). The tests of the WAIS fall into two broad categories, the verbal and the performance scale. Tests belonging to the verbal scale generally exhibit small age-related effects but tests belonging to the performance scale show a negative relationship with age (Matarazzo, 1976).

To date, the Seattle Longitudinal Study (Schaie, 1983) provides the most comprehensive picture of adult age gradients in psychometric intelligence in the United States. Since 1956, Schaie and his associates have been administering the PMA battery of primary mental abilities (Thurstone & Thurstone, 1949) as well as other tests and questionnaires in seven-year intervals to large longitudinal and cross-sectional samples

of adults of different ages and birth cohorts. Figure 1, based on the fifth data collection wave, displays cross-sectional adult age gradients based on multiple indicators for six intellectual abilities (Schaie & Willis, 1993). The next section has a comparison with longitudinal age changes. In essence, the age gradients closely resemble those obtained with the WAIS and the Army Alpha: Verbal ability and number ability peak during middle adulthood and show little or no age decrements before the age of 74, whereas perceptual speed, inductive reasoning, spatial orientation, and verbal memory show steady monotonic decline.

The simultaneous existence of stability and decline is also evident at the level of individuals and their change patterns. For instance, using across-person variability as a reference point, Schaie (1989) computed cumulative hazard functions to determine the age at which individual subjects experience statistically significant decline in one or more of five PMA ability tests. For the vast majority of individuals, at least two of the five primary mental abilities remained stable up to age 74 according to this criterion. In addition, some individuals evinced significant performance increments up into their seventies (Schaie, 1988).

CROSS-SECTIONAL VERSUS LONGITUDINAL DESIGNS: CONVERGENCE OF RESULTS

The area of psychometric intelligence was a domain where problems of age-related methodology were articulated and hotly debated (Baltes & Schaie, 1976; Horn & Donaldson, 1976). This debate has resulted in better methods and more generalizable evidence.

Information about adult age gradients in intellectual functioning has come from studies using cross-sectional, longitudinal, and mixed (e.g., cohort-sequential) sampling schemes. In cross-sectional studies, individuals of different ages are assessed at the same point in time. For this reason, cross-sectional age-group differences reflect not only age effects but also effects because of differences in birth cohort (generation). In longitudinal studies, individuals as they age are assessed repeatedly across time. Longitudinal designs are an indispensable tool of developmental research because they provide direct information about interindividual differences in intraindividual change

(Nesselrode, 1991). At the same time, the interpretability of average longitudinal age gradients is hampered by practice effects and selective attrition (i.e., longitudinal-study dropout and selective survival; Siegler & Botwinick, 1979). Thus, cross-sectional and longitudinal data are open to different sources of bias, and results obtained with the two sampling schemes have to be considered in combination to make optimal use of the available information.

A first step toward the joint interpretation of cross-sectional and longitudinal age gradients is to compare the performance of same-aged individuals across historical time (i.e., time-lagged comparisons). With some exceptions (e.g., number ability; cf. Schaie, 1989), the general picture resulting from these comparisons is that higher test scores are obtained at more recent times (Flynn, 1984; Schaie, 1983). Probably, this historical increase in test scores for same-aged individuals across historical time is not due to changes in the genetic composition of the population or differential sampling bias, but reflects some general change (improvement) in health- and education-related conditions.

Studies with cohort-sequential (e.g., mixed) designs such as the Seattle Longitudinal Study (Schaie, 1983) are well suited for three different kinds of comparisons across age: cross-sectional, longitudinal, and independent-sample same-cohort comparisons (e.g., age comparisons based on independent samples from the same birth cohort). The joint consideration of these three sampling schemes allows for better estimates of average age gradients. With respect to the Seattle Longitudinal Study, for example, independent-sample same-cohort and cross-sectional comparisons yielded practically identical estimates of seven-year change after controlling for the general increase in performance over historical time revealed by time-lagged comparisons (Salthouse, 1991b; Schaie, 1983, 1994). In contrast, longitudinal age changes, also corrected for historical change, showed somewhat less of a decrement with age. Given the convergence between cross-sectional and independent-sample same-cohort comparisons, the more positive age gradients found with longitudinal samples may partly derive from practice effects and selective attrition.

This analysis illustrates that in Western samples of the twentieth century, cross-sectional comparisons

may not necessarily lead to a distorted (i.e., much too negative) picture of adult age gradients in psychometric intelligence. Similar conclusions were reached by McArdle (cited in Horn & Hofer, 1992), who used structural modeling techniques to analyze simultaneously longitudinal and cross-sectional data obtained with the WAIS, and by Hertzog and Schaie (1988) in their structural modeling reanalysis of the Seattle data.

PREDICTIVE CONTINUITY VERSUS DISCONTINUITY OF INTELLECTUAL FUNCTIONING IN OLD AGE

Age-related individual differences in old age intellectual functioning are often seen as the cumulative outcome of a life history of cultural-environmental-educational conditions interacting with genetic factors. The enduring existence of this ensemble of conjoint influences is at the basis of the continuity view of interindividual differences in intellectual functioning. Generally, longitudinal data from middle adulthood and early old age support this view. For instance, Hertzog and Schaie (1986) found that seven-year stability coefficients (reliability-adjusted) for a general ability composite of the PMA battery ranged from .89 to .96 in samples with mean ages between 25 and 67 years at first test. In addition to prior levels of intellectual functioning, other factors, such as years of education, work complexity, and a life-history of hypertensive/cardiovascular symptomatology also contribute to predictive continuity. A major portion of the variance in these enduring sources of interindividual variability appears to be related to genetic differences (Pedersen et al., 1992; Plomin & Bergeman, 1991).

Without doubt, the most important health-related reason for discontinuity in level and rank order is the incidence of a dementing illness (Elias, Elias, & Elias, 1990). Dementia is a syndrome characterized by a global impairment of intellectual capacities such as memory, judgment, and orientation. The dominant pathology leading to dementia is Alzheimer's disease, which is strongly related to chronological age and shows major increases in prevalence rates beginning in the seventies.

More data about other potential sources of discontinuity, such as sensory functioning, health status, and concurrent activity level, are needed to determine

whether old and very old age, even in the absence of a dementing illness, is a relatively "autonomous" (i.e., life-history independent) phase of adult intellectual development (cf. Rabbitt, 1990). There is emerging evidence for this possibility. Lindenberger and Baltes (1994), for example, found an exceedingly strong connection between visual and auditory acuity and intellectual functioning in very old age and have argued that sensory functioning may emerge as a new correlate or antecedent of intellectual functioning in old age.

In general, present longitudinal evidence may overemphasize predictive continuity. For instance, the common practice to restrict final analyses of longitudinal data to survivors probably introduces a bias toward high predictive continuity because individuals with a debilitating or terminal illness are more likely to drop out of the sample than well-functioning individuals. Recent advances in structural modeling techniques with nonrandom missing data (McArdle et al., 1991) may help to overcome this bias.

THREE GENERALIZATIONS FROM THE PSYCHOMETRIC LITERATURE

The psychometric data on adult age gradients in intellectual functioning can be summarized in three points (Botwinick, 1977; Salthouse, 1991b). First, age trends in intellectual abilities are gradual rather than abrupt; on the group level, there is little evidence for sudden jumps or stepwise functions. Second, at least up to the seventies, the amount of variability attributable to chronological age is relatively small compared to the total amount of interindividual variability. For instance, the means of adults in their sixties are generally within one or two standard deviations of the distributions of adults in their early twenties (Figure 1). Age-ability correlations over the same age range rarely exceed $-.45$, which means that negative linear age gradients generally account for less than 20 percent of the total amount of interindividual variability in performance. Thus, at least for the period of "early" old age, age is not a powerful predictor of intellectual functioning for many practical purposes (Schaie, 1988).

The third generalization concerns the existence of multiple age gradients, that is, the distinction between vulnerable and maintained abilities (Horn & Hofer,

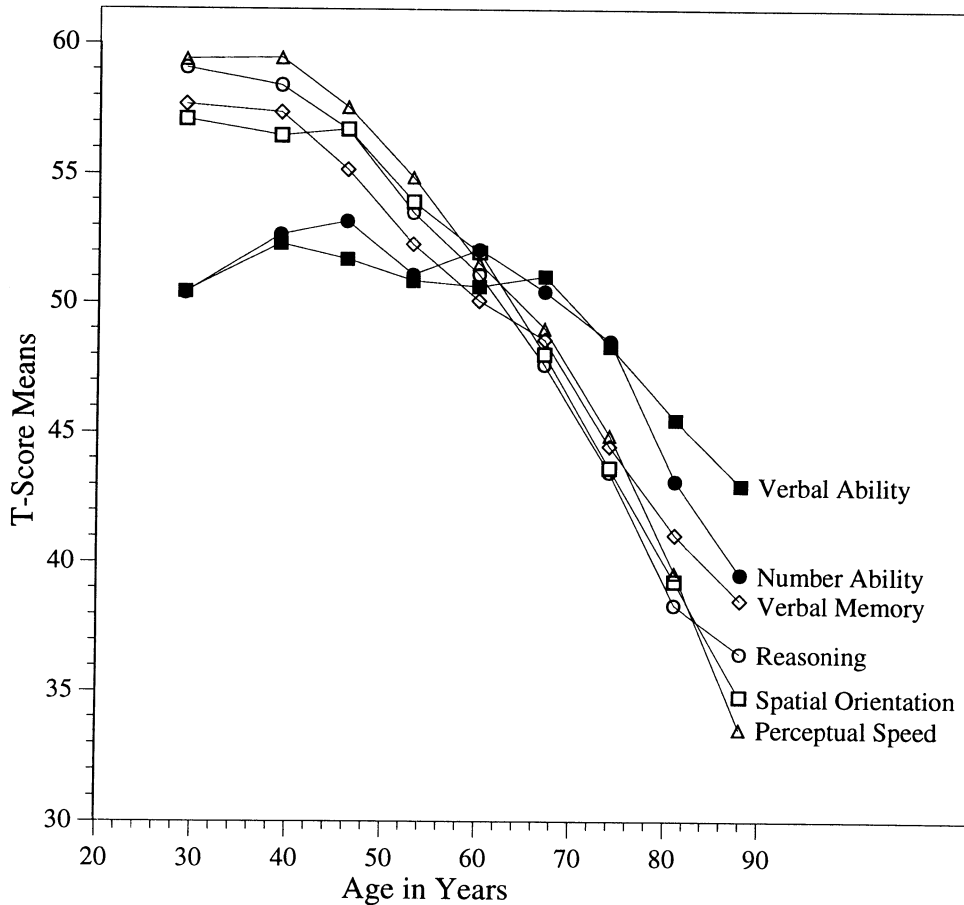


Figure 1

Cross-sectional age gradients in six primary mental abilities ($N = 1628$). Abilities were assessed with 3 to 4 different tests and were scaled in a T-score metric (i.e., mean = 50, SD = 10).

SOURCE: Based on Schaie & Willis (1993).

1992). Abilities such as reasoning, memory, spatial orientation, and cognitive/perceptual speed generally show a pattern of monotonic and roughly linear decline beginning in middle adulthood. In contrast, abilities such as verbal ability and number ability remain stable or increase up to the sixth or seventh decade of life.

ONTOGENETIC MODELS OF BIOLOGY-BASED VERSUS CULTURE-BASED INTELLIGENCE

The distinction between vulnerable and maintained abilities, which has been noticed for a long time (Hollingworth, 1927; Jones & Conrad, 1933), is reflected

in ontogenetic life-span theories of intelligence that contrast biological and cultural dimensions of cognitive functioning. Most prominent is the psychometric theory of fluid and crystallized intelligence (*Gf-Gc* theory; Cattell, 1971; Horn, 1982; see FLUID AND CRYSTALLIZED INTELLIGENCE, THEORY OF). According to *Gf-Gc* theory, intellectual abilities in the fluid domain reflect an individual's capacity to solve novel problems, to organize information, to ignore irrelevancies, to concentrate, and to maintain and divide attention. In contrast, crystallized abilities primarily reflect the acquisition and use of culturally valued bodies of knowledge. The distinction is similar to those of Donald HEBB between intellectual power and intellectual products and of Baltes's (1987, 1993) juxtaposition of the mechanics and pragmatics of cognition.

To explain the existence of different age gradients for fluid (i.e., vulnerable) and crystallized (i.e., maintained) abilities, *Gf–Gc* theory makes two basic assumptions. First, the theory posits that there is a gradual loss in brain efficiency with age during the adult life span. Second, it assumes that knowledge-based abilities are more resilient to this loss. To motivate the latter idea, Horn and Hofer (1992) invoked the notion of “overdetermination” in the context of neural networks. They assume that the kind of knowledge typical for crystallized abilities is implemented in highly interconnected networks that contain many possible ways to access a given piece of information (e.g., overdetermined information access). An age-associated loss in brain efficiency (connectivity) can be compensated as long as the number of remaining connections is sufficient to activate the relevant information.

Baltes (1993) combines the Horn-Cattell framework with perspectives from cognitive and evolutionary psychology. He links the fluidlike “mechanics” of cognition to the neurophysiological architecture of the mind as it developed during biological evolution (cf. Barkow, Cosmides, & Tooby, 1992) and associates the crystallizedlike “pragmatics” of cognition with the bodies of knowledge available from and mediated through culture.

The resulting theory of cognitive aging makes more explicit the need to distinguish between biological and social-cultural principles of “inheritance” and “transmission” (Durham, 1990). In the mechanics of the intellect, biological conditions reign supreme, and decline with aging is likely. In the pragmatics of the mind, however, the power of human agency and culture unfolds and, therefore, some progress may be possible into old age.

In comparison to *Gf–Gc* theory, the mechanic–pragmatic conception is conceptually broader in scope. For instance, the psychometric focus on sources of variability in intellectual functioning is supplemented with an interest in conditions affecting its level and range. Furthermore, the pragmatics of cognition embrace knowledge-based forms of intelligence better understood by theories of knowledge acquisition and expertise (Ericsson & Smith, 1991) than by extant psychometric approaches. Examples range from professional skills to knowledge regarding the meaning and conduct of life (e.g., wisdom).

PLASTICITY IN THE MECHANICS OF COGNITION: POTENTIAL AND LIMITS

Until the 1970s, a predominant position in psychometric intelligence research was to view performance potential on tests of intelligence as fairly fixed and immutable. As a consequence, negative age gradients in fluid ability test scores were often equated with irreversible decline (cf. Woodruff-Pak, 1989). With the advent of cognitive training research, the role of practice and of experiential factors in general became more central, and a research focus on modifiability and plasticity evolved (Baltes, 1993). Similar to developments in the field of trait psychology, research on the aging of psychometric intelligence increasingly turned to the search for contextual and situated conditions of intellectual dispositions (Berg & Sternberg, 1985; Sternberg & Wagner, 1986). (See PSYCHOMETRIC THEORIES OF INTELLIGENCE.)

Preserved Differentiation Versus Differential Preservation. A first approach to the issue of cognitive plasticity is to examine whether aging losses in cognitive performance are less likely to occur in individuals who lead an intellectually active life, either in general or with respect to specific abilities and skills. Here, the evidence is mixed. Contrary to what one might expect, little evidence suggests that age gradients on standard measures of psychometric intelligence vary substantially as a function of general intelligence, education, occupational status, or task-relevant experience (Salthouse, 1991b). In most cases, results are more consistent with what Salthouse and associates (1990) termed *preserved differentiation* (i.e., initial individual differences are maintained throughout adulthood) than with Denney’s (1984) perspective of differential preservation (i.e., practiced abilities do not decline). The longitudinal work by Kohn and Schooler (1983) on the relationship between the substantive complexity of work and ideational flexibility is a notable exception to this rule. Kohn and Schooler found that work complexity predicts increments in ideational flexibility over a period of ten years even after controlling for initial differences in ideational flexibility. However, the size of this effect was relatively small.

Activation of Learning Potential Among Older Adults. Intervention work (Baltes & Willis,

1982; Denney, 1984; Willis, 1987) is a more direct (i.e., experimentally controlled) way to explore the degree of plasticity in intellectual functioning. In addition, it examines whether age-related decrements in performance on intellectual ability measures are reversible, in full or in part, through training and practice (Schaie & Willis, 1986; Willis & C. Nesselrode, 1990). For the most part, interventions involved older adults only, focused on tests from the fluid ability domain, and used one of several intervention strategies such as tutored practice in task-related problem-solving skills, self-guided practice, or training in personality-related performance conditions such as self-efficacy beliefs.

The major results of this cognitive intervention work can be summarized in five points (cf. Baltes & Lindenberger, 1988):

1. Training gains in the practiced tests among healthy older adults are substantial (i.e., they roughly correspond to the amount of “naturally” occurring longitudinal decline between 60 and 80 years of age.)
2. Transfer, however, is limited to similar tests of the same ability.
3. Training gains are maintained over lengthy periods of time up to several years (Willis & C. Nesselrode, 1990).
4. The factor structure of the ability space is not altered substantially through training (Schaie et al., 1987).
5. Only in persons at risk for Alzheimer’s disease or afflicted by other forms of brain disease, training gains have been found to be severely reduced or nonexistent (M. Baltes, Kühl, & Sowarka, 1992).

These results indicate that the majority of “healthy” older adults, including those who display the typical pattern of age-related losses in fluid abilities under untrained conditions, are able to greatly improve their performance on fluid ability tests after a few sessions of task-related training or practice. Thus, cognitive plasticity in the fluid mechanics of cognition is preserved into old age and easily activated through experiential manipulations. However, there is little evidence so far to suggest that these training gains generalize to related abilities or to everyday functioning.

Age Comparisons in Maximum Levels of Performance in the Cognitive Mechanics.

What about age differences in cognitive plasticity? In analogy to “stress tests” used in biology and medicine (Fries & Crapo, 1981), “testing the limits” (Baltes, 1987; Kliegl & Baltes, 1987) was introduced as a paradigm for life-span research to assess age differences in maximum levels of performance by practice and variations in task difficulty. In this paradigm, large amounts of practice and experience under facilitative and supportive conditions are provided to gain insight into the latent reserves of individuals of different ages. Of special interest is the question whether young and older adults differ in the asymptotic levels of their “best” performance.

Age-comparative testing-the-limits research has been conducted most extensively with tasks involving memory for new information (Verhaegen, Marcoen, & Goossens, 1992). For instance, young and older adults were instructed and trained in the use of a mnemonic technique, the Method of Loci. The key feature of this method is to acquire a mental map of fixed locations, and to create mental images for each word that link the word to be remembered to one of the locations of the list. With this memory technique, subjects are able to recall long word lists—20, 30 words and more—in correct serial order after a single presentation. High correlations between performance in the Method of Loci and measures of fluid intelligence indicate that the method assesses individual differences in the fluid mechanics (Kliegl, Smith, & Baltes, 1990).

Figure 2 summarizes the results of a typical age-comparative experiment using a testing-the-limits approach (Baltes & Kliegl, 1992). The study involved a total of thirty-eight sessions of training and practice in the Method of Loci distributed over one year. Subjects were positively selected in terms of health and education to minimize pathological aging and cohort effects. Two findings are noteworthy. First, adults in both age groups greatly improved their memory performance, which demonstrates the continued existence of cognitive plasticity in old age. Second, however, practice and training resulted in a close-to-perfect separation of the two age groups and the demonstration of sizeable negative age differences at limits of functioning. Even after thirty-eight sessions of training, most older adults did not reach the level of per-

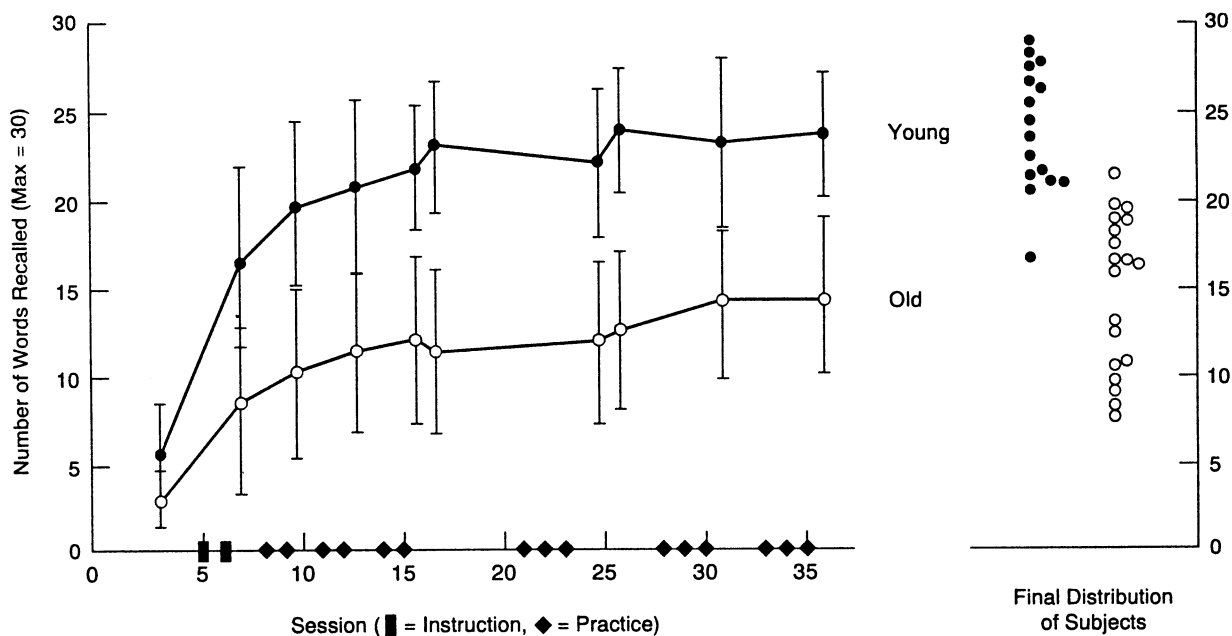


Figure 2

Testing-the-limits research suggests the existence of robust age-related losses in the mechanics of cognition. The example given involves a memory technique, the Method of Loci. After 38 sessions of training, most older adults did not reach the level of performance reached by young adults after only a few sessions. In the final distribution, not a single older person functioned above the mean of young adults.

SOURCE: Adapted from Baltes and Kliegl (1992).

formance that young adults had reached after only a few sessions. Moreover, at the end of the study, not a single older person functioned above the mean of young adults. In a subsequent study (Lindenberger, Kliegl, & Baltes, 1992), negative age differences of similar magnitude were found for older adults who had been selected for their presumed “talent” and job-related experience in one of the cognitive operations (i.e., mental imagery) relevant for the use of the memory technique.

THE SEARCH FOR AGING-RELATED PROCESSING CONSTRAINTS

The research reported so far has demonstrated that a major part of cognitive aging is decline, albeit with substantial variations in rate and onset by abilities and persons. Not surprisingly, therefore, there have been attempts to identify possible sources for negative age-related decrements in cognitive functioning. Specifi-

cally, some researchers (Birren, 1964; Cerella, 1990; Salthouse, 1991b) have initiated a search for age-associated domain-general constraints that would fuel negative adult age gradients in the mechanics of cognition. Ideally, the identification of such constraints or limiting resources would permit a unified and parsimonious explanation of negative age differences in cognition by mapping age gradients in a variety of cognitive performances onto a small set of domain-general constraining parameters. Furthermore, the identification of age-based general processing constraints may help to explain life-span changes in the mean and covariance structure of intelligence.

The perils in the search for general age-related constraints on information processing are that cognitive aging phenomena are rephrased in resource terminology rather than explained, and that evidence in favor of process-specificity is overlooked (for critical discussions, see Hasher & Zacks, 1988; Light, 1991). To minimize these risks, hypothesized developmental con-

straints need to be specified with sufficient precision to make contact with cognitive processes and to permit differential predictions (Mayr & Kliegl, 1993). The following section reports evidence related to three concepts currently discussed as possible candidates for aging-related general processing constraints in the fluid mechanics of cognition: speed, working memory capacity, and inhibition. This list is not exhaustive but serves to illustrate the general approach.

Speed of Information Processing. Researchers of cognitive aging generally agree that performance on measures assessing cognitive or perceptual processes slows down with advancing age. Proponents of the general slowing hypothesis of cognitive aging argue that this phenomenon is caused by a general decrease in processing rate with age (Birren, 1964; Welford, 1984). Evidence supporting the assumption of generalized forms of slowing has come from several sources including meta-analyses of latency data examining the relation between mean latencies for groups of old adults and the mean latencies for groups of young adults (Cerella, 1990; Myerson et al., 1990).

Recent psychometric evidence has supported the prediction that speed is a major factor in producing or mediating negative adult age differences in other intellectual abilities, even if these other abilities are assessed under time-relaxed or untimed testing conditions (Hertzog, 1989; Lindenberger, Mayr, & Kliegl, 1993; Salthouse, 1991a). For example, negative age gradients in fluid abilities such as reasoning or short-term memory were greatly reduced or eliminated, and age gradients in crystallized abilities remained positive up to age 70 after statistically controlling for cognitive/perceptual speed (Hertzog, 1989).

It needs to be acknowledged, however, that the explanatory construct "speed" is neither fully understood nor independently established. Extant psychometric measures of cognitive/perceptual speed probably involve a relatively complex sequence of components such as sensory processes, working memory, secondary memory, and motor functioning. Age-comparative research using more direct measures of information processing rate such as nerve conduction velocity (Vernon & Mori, 1992) is needed to further test the viability of the speed hypothesis.

In this context, it also appears fruitful to revisit the question of differential age effects in speed versus

power conditions of testing (Botwinick, 1977). This issue can be studied systematically by investigating age differences in the function relating presentation time to accuracy level. For example, the asymptotes of time-accuracy functions may differ by age, suggesting that certain cognitive operations are beyond reach for older individuals no matter how much presentation time is available.

Working Memory. Negative age differences in working memory (Baddeley, 1992) have also been invoked as a possible cause for negative age gradients in adult intelligence (Craik, 1983; Craik & Byrd, 1982). Working memory denotes the ability to preserve information in a short-term store while simultaneously transforming the same or some other information.

One way to study the effect of age differences in working memory on intellectual functioning is to vary the relative importance of temporary storage and processing (i.e., information transformation) demands within or across tasks. With few exceptions, this research has demonstrated that age differences are more pronounced when demands on processing are increased (Mayr & Kliegl, 1993; Craik & Jennings, 1992, for a critical discussion).

Another way to investigate the importance of working memory for adult age differences in cognitive functioning is to examine whether separately administered working memory measures predict age differences in other measures of intellectual functioning. Studies of that sort indicate that working memory measures contribute to age differences in a variety of fluid intelligence tasks, and that a significant portion of the predictive variance in the working memory measures is orthogonal to cognitive/perceptual speed (Salthouse, 1991a).

Age-Related Decline in the Efficiency of Inhibitory Processes. Adult age differences in intellectual functioning and cognitive style may also reflect a reduction in the efficiency of inhibitory processes (Hasher & Zacks, 1988; cf. Pascual-Leone, 1983). To process goal-relevant information efficiently, working memory is aided by inhibitory mechanisms that shelter working memory against goal-irrelevant information. Age-related decrements in the efficiency of these inhibitory mechanisms would permit a larger amount of irrelevant information to enter working memory and to receive sustained attention.

The inhibition hypothesis is consistent with findings from different strands of cognitive aging research. For instance, large negative age differences have been found for measures that require selective attention and the inhibition of goal-irrelevant information (Dempster, 1992). Moreover, older adults were found to show more proactive interference in cued recall of word lists (Kliegl & Lindenberger, 1993). The most direct support for the inhibition hypothesis has come from studies on negative priming where older adults were found to be less slowed than younger adults in responding to a stimulus on a current trial that served as a distractor on the previous trial (e.g., Hasher et al., 1991). Taken together, these results are promising and call for more research to examine the relative importance of age differences in inhibitory mechanisms for age gradients in intellectual functioning.

THE “CRYSTALLIZED” PRAGMATICS OF COGNITION: EVIDENCE FOR STABILITY AND GROWTH

The distinction between “biological” and “cultural” intelligence also directs attention toward facets of cognitive functioning that exhibit potential for continued growth in adulthood and old age (Baltes, 1993; Perlmutter, 1990). Growth and stability are possible, if not expected, whenever intellectual performance is knowledge-driven and does not require exceedingly high accuracy or speed in mechanical functioning.

Under the heading of crystallized intelligence, psychometric research (Horn & Hofer, 1992) has identified instances of knowledge-based stability and growth in skills that are widely distributed in a given culture and amenable to psychometric testing (e.g., verbal ability, number ability). However, other skills and domains of knowledge have escaped psychometric operationalization. To obtain a more comprehensive picture of the cognitive pragmatics, the psychometric tradition needs to be supplemented by approaches with a more explicit focus on knowledge acquisition and utilization.

The expertise approach appears especially fruitful in this regard. Originally, the study of expertise has been restricted to the componential analysis of skills in the laboratory, chess being the most prominent example (Simon & Chase, 1973). Since then, the expertise approach has been applied to domains of broader

developmental significance including the concept of wisdom.

Wisdom: Expertise in the Fundamental Pragmatics of Life. Most people, lay persons and researchers alike, regard wisdom as the hallmark of positive intellectual aging (Sternberg, 1990). According to expertise-oriented conceptions of WISDOM, high levels of wisdom-related knowledge and skills are assumed to reflect a combination of personality dispositions, extended practice in matters regarding the human condition, and the presence of effective mentors. Age may be positively related to the joint occurrence of these wisdom-enhancing factors because their operation requires experience in diverse conditions of life and because certain facets of adult personality evolving in advanced age such as generativity and integrity may promote development in the direction of wisdom (e.g., Blanchard-Fields, 1989; Sternberg, 1990).

Based on these considerations, a research group at the Berlin Max Planck Institute (Baltes, Smith, & Staudinger, 1992; Baltes & Staudinger, 1993) has defined wisdom as an expert knowledge system in the “fundamental pragmatics of life permitting excellent judgment and advice involving important and uncertain matters of life.” In a first series of studies, thinking-aloud responses to difficult life problems were scored on five wisdom-related criteria (factual knowledge, procedural knowledge, contextualism, relativism, and uncertainty). Two findings are noteworthy, both of which stand in sharp contrast to findings related to the aging of the fluid mechanics. First, when comparing adults of about 30 to 70 years of age, there were no major age differences in average performance. Second, older persons with wisdom-facilitative experiences (e.g., older clinical psychologists) contributed a disproportionately large share to the top responses.

Knowledge About Emotions and the Self. Another area of potential growth in the pragmatics of cognition concerns the interplay among cognition, emotion, and the self (Standing, Marsiske, & Baltes, 1993). Labouvie-Vief, DeVoe, and Bulka (1990) examined conceptions about emotions in a sample of 72 individuals aged 10 to 77 years. Subjects’ self-descriptions of several emotions were rated in accordance to a four-level model of emotion regulation. Adolescents tended to describe emotions in terms of sensorimotor

action, outer appearance, conventional descriptions, and rigid impulse monitoring, whereas adults conveyed a more vivid sense of the experience, possessed more explicit knowledge of bodily sensations, tended to accept conflict within self and others, and displayed more flexibility and delay of action.

Viewed from an expertise perspective, these results suggest that aging individuals become more knowledgeable about their own emotions, feelings, and intentions. This knowledge-based interpretation is consistent with other recent approaches focusing on positive aspects of aging such as socioemotional selectivity theory (Carstensen, 1993) and the theory of selective optimization with compensation (Baltes & Baltes, 1990). Moreover, it resonates with recent accounts of young children's development of their knowledge about their own mental states (Gopnik, 1993).

THE INTERPLAY BETWEEN GROWTH AND DECLINE: COMBINING EXPERTISE AND INFORMATION-PROCESSING APPROACHES

The evidence presented so far has identified two sides of cognitive aging: losses in the fluid cognitive mechanics and the potential for stability and growth in those domains of functioning where knowledge is an important ingredient. A new line of research addresses explicitly the interaction between these two aspects of functioning. Such research is nurtured by the notions of expertise (Knopf, Kolodziej, & Preussler, 1990) and compensation (Bäckman & Dixon, 1992; Baltes & Baltes, 1990), and has yielded new insights into the nature of cognitive aging.

Molar Equivalence/Molecular Decomposition. Both deliberate practice, a key characteristic of expertise in the narrow sense (Ericsson, Krampe, & Tesch-Römer, 1993), and aging-induced reductions in processing efficiency take place over extended periods of time (e.g., decades). One goal in cognitive aging research is to understand the resulting coevolution of decreasingly effective general-purpose mechanisms and increasing domain-specific skills. In the molar equivalence/molecular decomposition paradigm (Salthouse, 1984), adults of different ages are equated in criterion task proficiency to investigate whether equal

levels of criterion performance are attained through age-differential profiles of component processes (Charness, 1989).

For example, Salthouse (1984) studied a total of 74 transcription typists ranging from 19 to 72 years of age. Age and skill level (i.e., net words per minute) were uncorrelated. Age was negatively related to measures of perceptual/motor speed (e.g., tapping speed) but positively related to eye-hand span. In other words, older typists were slower but looked further ahead in the text to be typed. These findings are consistent with the interpretation that aging typists extend their eye-hand span to counteract the consequences of aging losses in perceptual/motor speed (Figure 3), and illustrate the compensatory relationship between knowledge and speed.

Age Differences in Peak Performance. Experts are often older than novices, presumably because it takes prolonged periods of deliberate practice to reach excellence in a particular domain of functioning. Lehman (1953) found that peak performances in the arts and sciences are most frequently seen in the third to fifth decade of life. The precise age at which peak performances are most likely to occur apparently depends on the life course of opportunity structures as well as on the relative importance of knowledge and basic processing efficiency in a given domain.

A good example comes from chess (Charness & Bosman, 1990). The mean age at which a world championship is first won is about 46 years of age for correspondence chess, but about 30 years of age for tournament chess. In correspondence chess, players are permitted three days to deliberate a move; in tournament chess, deliberation averages three minutes per move. The difference in peak age between the two activities seems to reflect differences in the relative importance of cognitive/perceptual speed and knowledge.

Creativity and Coping with Death: The "Swan-Song" Phenomenon. Some evidence shows that certain types of peak performance occur very late in life. As expected by expertise-based views of exceptional performance, individuals who start their careers early and produce at extraordinary rates are the ones who are most likely to remain creative in their last years of life (Simonton, 1988). One example is Sophocles (497–406 B.C.), who won his first prize for the

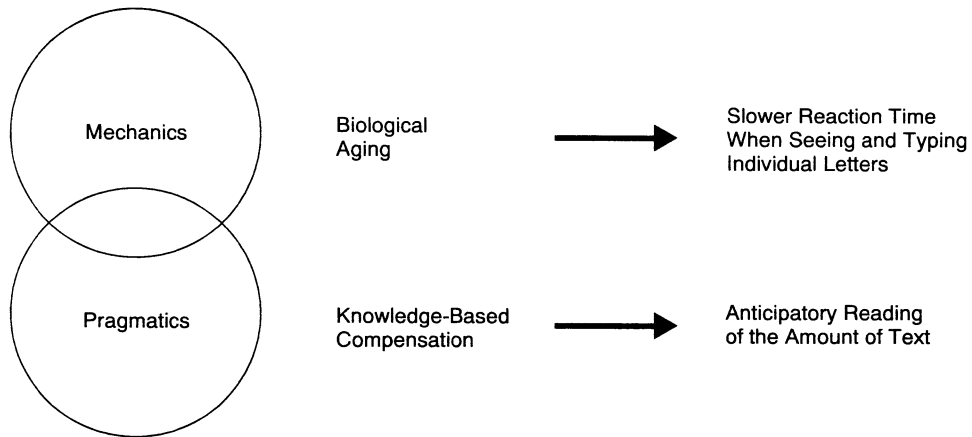


Figure 3

Older typists can maintain a high level of functioning by reading further ahead in the text to be typed, despite a loss in reaction time when typing individual letters.

SOURCE: Based on Salthouse (1984).

best drama of the year at the age of 28, wrote over 120 dramas, and developed a new dramatic style in his eighties. Commenting on his own late-life artistic development, Sophocles said that he finally had liberated himself from the artificiality of his earlier style and had found a language that was “the best and the most ethical” (Schadewaldt, 1975, p. 75).

Recent evidence on classical composers has substantiated the claim that special forms of late-life creativity may in fact exist. Simonton (1989) proposed that artists facing death may feel the need to make optimal use of their limited future. He examined this issue by assessing the relationship between closeness to death and a set of criterion variables for a sample of 1,910 works written by 172 classical composers. Last works scored lower in melodic variability and performance duration, but higher in repertoire popularity and aesthetic significance. The tendency toward condensed expression of the essential may be a general feature of late-life creativity.

CONCLUSION

Intellectual aging comprises decline, maintenance, and growth. Its multidirectionality and multidimensionality is captured by life-span models that distinguish two streams of inheritance and transmission, the biological (the fluidlike mechanics) and the cultural

(the crystallized pragmatics). In the fluid mechanics, negative age changes and differences prevail, even after controlling for generational cohort differences, historical changes in the amount of education, and age differences in the proportion of individuals with serious health problems. Negative age differences are especially pronounced at maximum levels of performance. Research on age differences in information processing constraints suggests that constructs such as processing speed, working memory capacity, and the ability to inhibit irrelevant information may capture important dimensions of mechanic decline.

However, the notion of decrements in the fluid mechanics needs to be qualified by two statements. First, the average level of functioning reached by current aging cohorts is not fixed. Rather, cognitive training work suggests that higher levels are attainable by adults of all ages including the very old. Second, standard fluid ability measures generally show a substantial amount of overlap between young and old adults.

In contrast to the mechanics, the knowledge-saturated pragmatics of cognition are often associated with maintenance and growth. In the psychometric domain, crystallized intellectual abilities such as verbal knowledge increase or remain stable up to age 60 and beyond. Work informed by the study of expertise has accumulated new evidence in favor of a life-span conception of intelligence that captures the potential for

stability and growth associated with the acquisition and refinement of domain-specific knowledge and skills. For example, when combined with facilitating environmental and person-related conditions, growing old can be associated with higher levels of self- and wisdom-related performance. Finally, the combined application of expertise and information-processing approaches has helped to clarify how the adverse effects of cognitive decline in the mechanics can be compensated through specialized bodies of knowledge.

(See also: AGE AND THE CONTENT OF INTELLIGENCE TESTS.)

BIBLIOGRAPHY

- ALEXANDER, C., & LANGER, E. J. (EDS.). (1990). *Beyond formal operations: Alternative endpoints to human development*. New York: Oxford University Press.
- BÄCKMAN, L., & DIXON, R. A. (1992). Psychological compensation: A theoretical framework. *Psychological Bulletin*, *112*, 259–283.
- BADDELEY, A. D. (1992). Working memory. *Science*, *255*, 556–559.
- BALTES, M. M., KÜHL, K.-P., & SOWARKA, D. (1992). Testing the limits of cognitive reserve capacity: A promising strategy for early diagnosis of dementia? *Journal of Gerontology*, *47*, 165–167.
- BALTES, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, *23*, 611–626.
- BALTES, P. B. (1993). The aging mind: Potential and limits. *Gerontologist*, *33*, 580–594.
- BALTES, P. B., & BALTES, M. M. (1990). Psychological perspectives on successful aging: The model of selective optimization with compensation. In P. B. Baltes & M. M. Baltes (Eds.), *Successful aging: Perspectives from the behavioral sciences* (pp. 1–34). Cambridge: Cambridge University Press.
- BALTES, P. B., & KLIEGL, R. (1992). Further testing of limits of cognitive plasticity: Negative age differences in a mnemonic skill are robust. *Developmental Psychology*, *28*, 121–125.
- BALTES, P. B., & LABOUIVIE, G. V. (1973). Adult development of intellectual performance: Description, explanation, modification. In C. Eisdorfer & M. P. Lawton (Eds.), *The psychology of adult development and aging* (pp. 365–395). Washington, DC: American Psychological Association.
- BALTES, P. B., & LINDENBERGER, U. (1988). On the range of cognitive plasticity in old age as a function of experience: 15 years of intervention research. *Behavior Therapy*, *19*, 283–300.
- BALTES, P. B., & SCHAIE, K. W. (1976). On the plasticity of intelligence in adulthood and old age: Where Horn and Donaldson fail. *American Psychologist*, *31*, 720–725.
- BALTES, P. B., SMITH, J., & STAUDINGER, U. M. (1992). Wisdom and successful aging. In T. Sonderegger (Ed.), *Nebraska symposium on motivation* (Vol. 39, pp. 123–167). Lincoln: University of Nebraska Press.
- BALTES, P. B., & STAUDINGER, U. M. (1993). The search for a psychology of wisdom. *Current Directions in Psychological Science*, *2*, 75–80.
- BALTES, P. B., & WILLIS, S. L. (1982). Plasticity and enhancement of intellectual functioning in old age: Penn State's Adult Development and Enrichment Project (ADEPT). In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and cognitive processes* (pp. 353–389). New York: Plenum.
- BARKOW, J. H., COSMIDES, L., & TOOBY, J. (EDS.). (1990). *The adapted mind: Evolutionary psychology and the generation of culture*. New York: Oxford University Press.
- BERG, C. A., & STERNBERG, R. J. (1985). A triarchic theory of intellectual development during adulthood. *Developmental Review*, *5*, 334–370.
- BIRREN, J. E. (1964). *The psychology of aging*. Englewood Cliffs, NJ: Prentice-Hall.
- BLANCHARD-FIELDS, F. (1989). Postformal reasoning in a socioemotional context. In M. L. Commons, J. D. Sinnott, F. A. Richards, & C. Armon (Eds.), *Adult development: Comparisons and applications of developmental models* (pp. 77–93). New York: Praeger.
- BOTWINICK, J. (1977). Intellectual abilities. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (1st ed., pp. 580–605). New York: Van Nostrand Reinhold.
- CARSTENSEN, L. (1993). Motivation for social contact across the life span: A theory of socioemotional selectivity. In J. Jacobs (Ed.), *Nebraska symposium on motivation* (Vol. 40, pp. 205–254). Lincoln: University of Nebraska Press.
- CATTELL, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton-Mifflin.
- CERELLA, J. (1990). Aging and information-processing rate.

- In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 201–221). San Diego, CA: Academic Press.
- CHARNESS, N. (1989). Age and expertise: Responding to Taland's challenge. In L. W. Poon (Ed.), *Everyday cognition in adulthood and late life* (pp. 437–456). Cambridge: Cambridge University Press.
- CHARNESS, N., & BOSMAN, E. A. (1990). Expertise and aging: Life in the lab. In T. H. Hess (Ed.), *Aging and cognition: Knowledge organization and utilization* (pp. 343–385). Amsterdam: Elsevier.
- CRAIK, F. I. M. (1983). On the transfer of information from temporary to permanent memory. *Philosophical Transactions of the Royal Society of London, B* 302, 341–359.
- CRAIK, F. I. M., & BYRD, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191–211). New York: Plenum.
- CRAIK, F. I. M., & JENNINGS, J. M. (1992). Human memory. In F. I. M. Craik & T. A. Salthouse (Eds.), *The handbook of aging and cognition* (pp. 51–110). Hillsdale, NJ: Erlbaum.
- DEMPSTER, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review, 12*, 45–75.
- DENNEY, N. W. (1984). A model of cognitive development across the life span. *Developmental Review, 4*, 171–191.
- DIXON, R. A., & BALTES, P. B. (1986). Towards life-span research on the functions and pragmatics of intelligence. In R. J. Sternberg & R. K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world* (pp. 203–235). Cambridge: Cambridge University Press.
- DURHAM, W. H. (1990). Advances in evolutionary culture theory. *Annual Review of Anthropology, 19*, 187–210.
- ELIAS, M. F., ELIAS, J. W., & ELIAS, P. K. (1990). Biological and health influences on behavior. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 80–102). New York: Academic Press.
- ERICSSON, K. A., KRAMPE, R. T., & TESCH-RÖMER, C. (1993). The role of deliberate practice in the acquisition of expert performance. *Psychological Review, 100*, 363–406.
- ERICSSON, K. A., & SMITH, J. (EDS.). (1991). *Towards a general theory of expertise: Prospects and limits*. New York: Cambridge University Press.
- FLYNN, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin, 95*, 29–51.
- FRIES, J. F., & CRAPO, L. M. (1981). *Vitality and aging*. San Francisco: W. H. Freeman.
- GOPNIK, A. (1993). How we know our minds: The illusion of first-person knowledge of intentionality. *Behavioral and Brain Sciences, 16*, 1–14.
- HASHER, L., STOLTZFUS, E. R., ZACKS, R. T., & RYPMA, B. (1991). Age and inhibition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 163–169.
- HASHER, L., & ZACKS, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. *The Psychology of Learning and Motivation, 22*, 193–225.
- HEBB, D. O. (1949). *The organization of behavior*. New York: Wiley.
- HERTZOG, C. (1989). Influences of cognitive slowing on age differences in intelligence. *Developmental Psychology, 25*, 636–651.
- HERTZOG, C., & SCHAIE, K. W. (1986). Stability and change in adult intelligence: 1. Analysis of longitudinal covariance structures. *Psychology and Aging, 1*, 159–171.
- HERTZOG, C., & SCHAIE, K. W. (1988). Stability and change in adult intelligence: 2. Simultaneous analysis of longitudinal means and covariance structures. *Psychology and Aging, 3*, 122–130.
- HOLLINGWORTH, H. L. (1927). *Mental growth and decline*. New York: Appleton.
- HORN, J. L. (1982). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 237–278). New York: Plenum.
- HORN, J. L., & DONALDSON, G. (1976). On the myth of intellectual decline in adulthood. *American Psychologist, 31*, 701–709.
- HORN, J. L., & HOFER, S. M. (1992). Major abilities and development in the adult period. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 44–99). Cambridge: Cambridge University Press.
- JONES, H. E., & CONRAD, H. (1933). The growth and decline of intelligence: A study of a homogeneous group between the ages of ten and sixty. *Genetic Psychological Monographs, 13*, 223–298.
- KLIEGL, R., & BALTES, P. B. (1987). Theory-guided analysis of development and aging mechanisms through testing-the-limits and research on expertise. In C. Schooler & K. W. Schaie (Eds.), *Cognitive functioning and social structure over the life course* (pp. 95–119). Norwood, NJ: Ablex.

- KLIEGL, R., & LINDENBERGER, U. (1993). Modeling intrusions and correct recall in episodic memory: Adult age differences in encoding of list context. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 617–637.
- KLIEGL, R., SMITH, J., & BALTES, P. B. (1990). On the locus of magnification of age differences during mnemonic training. *Developmental Psychology*, *26*, 894–904.
- KNOF, M., KOLODZIEJ, P., & PREUSSLER, W. (1990). Der ältere Mensch als Experte—Literaturübersicht über die Rolle von Expertenwissen für die kognitive Leistungsfähigkeit im höheren Alter [Elderly experts—A review on the role of expertise in cognitive performance in older adults]. *Zeitschrift für Gerontopsychologie und Psychiatrie*, *4*, 233–248.
- KOHN, M. L., & SCHOOLER, C. (1983). *Work and personality*. Norwood, NJ: Ablex.
- LABOUVIE-VIEF, G. (1977). Adult cognitive development: In search of alternative interpretations. *Merrill Palmer Quarterly*, *23*, 277–263.
- LABOUVIE-VIEF, G. (1992). A neo-Piagetian perspective on adult cognitive development. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 197–228). Cambridge: Cambridge University Press.
- LABOUVIE-VIEF, G., DEVOE, M., & BULKA, D. (1989). Speaking about feelings: Conceptions of emotion across the life span. *Psychology and Aging*, *4*, 425–437.
- LEHMAN, H. C. (1953). *Age and achievement*. Princeton, NJ: Princeton University Press.
- LIGHT, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology*, *42*, 333–376.
- LINDENBERGER, U., & BALTES, P. B. (1994). Sensory functioning and intelligence in old age: A powerful connection. *Psychology and Aging*, *9*.
- LINDENBERGER, U., KLIEGL, R., & BALTES, P. B. (1992). Professional expertise does not eliminate age differences in imagery-based memory performance during adulthood. *Psychology and Aging*, *7*, 585–593.
- LINDENBERGER, U., MAYR, U., & KLIEGL, R. (1993). Speed and intelligence in old age. *Psychology and Aging*, *8*, 207–220.
- MATARAZZO, J. D. (1976). *Wechsler's measurement and appraisal of intelligence*. Baltimore: Williams & Wilkins.
- MCARDLE, J. J., HAMAGAMI, F., ELIAS, M. F., & ROBBINS, M. A. (1991). Structural modeling of mixed longitudinal and cross-sectional data. *Experimental Aging Research*, *17*, 29–51.
- MAYR, U., & KLIEGL, R. (1993). Sequential and coordinative complexity: Age-based processing limitations in figural transformations. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 1297–1320.
- MYERSON, J., HALE, S., WAGSTAFF, D., POON, L. W., & SMITH, G. A. (1990). The information-loss model: A mathematical theory of age-related cognitive slowing. *Psychological Review*, *97*, 475–487.
- NESSSELROADE, J. R. (1991). Interindividual differences in intraindividual change. In L. M. Collins & J. L. Horn (Eds.), *Best methods for the analysis of change: Recent advances, unanswered questions, future directions* (pp. 92–105). Washington, DC: American Psychological Association.
- PASCUAL-LEONE, J. (1983). Growing into human maturity: Towards a metasubjective theory of adulthood stages. In P. B. Baltes (Ed.), *Life-span development and behavior* (Vol. 5, pp. 117–156). New York: Academic Press.
- PEDERSEN, N. L., PLOMIN, R., NESSELROADE, J. R., & MCCLEARN, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science*, *3*, 346–353.
- PERLMUTTER, M. (ED.). (1990). *Late-life potential*. Washington, DC: Gerontological Society of America.
- PLOMIN, R., & BERGEMAN, C. S. (1991). The nature of nurture: Genetic influence on “environmental” measures. *Behavioral and Brain Sciences*, *14*, 373–386.
- RABBITT, P. (1990). Applied cognitive gerontology: Some problems, methodologies, and data. *Applied Cognitive Psychology*, *4*, 225–246.
- RYBASH, J. M., HOYER, W. J., & RODIN, P. A. (1986). *Adult cognition and aging: Developmental changes in processing, knowing, and thinking*. Elmsford, NY: Pergamon Press.
- SALTHOUSE, T. A. (1984). Effects of age and skill on typing. *Journal of Experimental Psychology: General*, *13*, 345–371.
- SALTHOUSE, T. A. (1991a). Mediation of adult age differences in cognition by reductions in working memory and speed of processing. *Psychological Science*, *2*, 179–183.
- SALTHOUSE, T. A. (1991b). *Theoretical perspectives on cognitive aging*. Hillsdale, NJ: Erlbaum.
- SALTHOUSE, T. A., BABCOCK, R. L., SKOVRONCK, E., MITCHELL, D. R. D., & PALMON, R. (1990). Age and experience effects in spatial visualization. *Developmental Psychology*, *26*, 128–136.
- SCHADEWALDT, W. (1975). *Nachwort zu Elektra [Epilogue to Elektra]*. Stuttgart: Reclam.
- SCHAIK, K. W. (1983). The Seattle Longitudinal Study: A

- 21-year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 64–135). New York: Guilford Press.
- SCHAE, K. W. (1988). Variability in cognitive functioning in the elderly: Implications for societal participation. In A. D. Woodhead, M. A. Bender, & R. C. Leonard (Eds.), *Phenotypic variation in populations: Relevance to risk assessment* (pp. 191–212). New York: Plenum.
- SCHAE, K. W. (1989). The hazards of cognitive aging. *The Gerontologist*, 29, 484–493.
- SCHAE, K. W. (1994). The course of adult intellectual development. *American Psychologist*, 49.
- SCHAE, K. W., & WILLIS, S. L. (1986). Can decline in adult intellectual functioning be reversed? *Developmental Psychology*, 22, 223–232.
- SCHAE, K. W., WILLIS, S. L., HERTZOG, C., & SCHULENBERG, J. E. (1987). Effects of cognitive training on primary mental ability structure. *Psychology and Aging*, 2, 233–242.
- SIEGLER, I. C., & BOTWINICK, J. (1979). A long-term longitudinal study of intellectual ability of older adults: The matter of selective subject attrition. *Journal of Gerontology*, 34, 242–245.
- SIMON, H. A., & CHASE, W. G. (1973). Skill in chess. *American Scientist*, 61, 394–403.
- SIMONTON, D. K. (1988). Age and outstanding achievement: What do we know after a century of research? *Psychological Bulletin*, 104, 251–267.
- SIMONTON, D. K. (1989). The swan-song phenomenon: Last-work effects for 172 classical composers. *Psychology and Aging*, 4, 42–47.
- STAUDINGER, U. M., MARSISKE, M., & BALTES, P. B. (1993). Resilience and levels of reserve capacity in later adulthood: Perspectives from life-span theory. *Development and Psychopathology*, 5, 541–566.
- STERNBERG, R. J. (ED.). (1990). *Wisdom: Its nature, origins, and development*. New York: Cambridge University Press.
- STERNBERG, R. J., & BERG, C. A. (EDS.). (1992). *Intellectual development*. New York: Cambridge University Press.
- STERNBERG, R. J., & WAGNER, R. K. (1986) (Eds.). *Practical intelligence: Nature and origins of competence in the everyday world*. Cambridge: Cambridge University Press.
- THURSTONE, L. L., & THURSTONE, T. G. (1949). *Examiner's manual: SRA Primary Mental Abilities Test (Form 11–17)*. Chicago: Science Research Associates.
- VERHAEGEN, P., MARCOEN, A., & GOOSSENS, L. (1992). Improving memory performance in the aged through mnemonic training: A meta-analytic study. *Psychology and Aging*, 7, 242–251.
- VERNON, P. A., & MORI, M. (1992). Intelligence, reaction time, and peripheral nerve conduction velocity. *Intelligence*, 16, 273–288.
- WELFORD, A. T. (1984). Between bodily changes and performance: Some possible reasons for slowing with age. *Experimental Aging Research*, 10, 73–88.
- WILLIS, S. L. (1987). Cognitive training and everyday competence. In K. W. Schaie (Ed.), *Annual review of gerontology and geriatrics* (Vol. 7, pp. 159–188). New York: Springer-Verlag.
- WILLIS, S. L., & NESSELROADE, C. S. (1990). Long-term effects of fluid ability training in old-old age. *Developmental Psychology*, 26, 905–910.
- WOODRUFF-PAK, D. S. (1989). Aging and intelligence: Changing perspectives in the twentieth century. *Journal of Aging Studies*, 3, 91–118.

ULMAN LINDENBERGER

PAUL B. BALTES

AGNOSIA Imagine waking up one morning and hearing familiar people talking but being unable to understand a word that is said, or seeing an object but having no idea what it is until it is placed in your hand and you immediately recognize it as your hairbrush. The term *agnosia* is defined as the inability to recognize sensory stimuli despite intact sensation (Kolb & Whishaw, 1985). A patient may demonstrate intact perception by accurately copying very detailed drawings, yet have no idea of what he or she has drawn. Another patient may orient to sound and know when people are talking, but his or her native tongue now sounds like a foreign language. With few exceptions, an agnosia represents the loss of an ability to recognize stimuli, rather than a developmental disorder. The definition of agnosia is not controversial; however, the mechanisms, types, and subdivisions within the agnosias are.

Any review of the literature on agnosia is plagued by confusion generated by differing definitions of terms and disagreement regarding when an ability is “intact.” Overlap between the classical neurologic disorders also adds to the confusion. If an inability to name is an aphasia and an inability to recognize is an

agnosia, is a better term for the inability to name and recognize a visually presented object *optic aphasia* or *visual object agnosia*? Most descriptions in this article will include alternate definitions and terms that may have been popular or suggested historically.

Determining whether someone can see, hear, or feel may seem simple, but deficits following brain injury are rarely all-or-nothing events. Sight, hearing, and touch may be partially or mostly preserved and yet not quite normal. A patient will be diagnosed with agnosia, not just a perceptual deficit, if other patients with a similar degree of impairment in perception are still able to recognize the same stimuli. Patients with these disorders rarely present themselves complaining of an agnosia, but will instead say that they “cannot hear right” or complain that there is something wrong with their eyes and that they need new glasses.

Mesulam (1985) provides a systematic description for diagnosing agnosia. Although specifically writing about visual agnosia, his points are equally valid for other types of agnosia. First, since language should be unimpaired in agnosia, a patient should be able to describe attributes of the object. Mesulam (1985) notes that “fine disturbances in perception may not be detectable” but may contribute to problems with recognizing an object. Although patients may behave as if they perceive the object, all brain areas normally involved in perception may not have access to that information (i.e., there may be a “disconnection” between important brain areas). Patients who are able to demonstrate intact perception must then be unable to recognize what they are perceiving. This must be separated from a disorder in naming. Patients who do not recognize something cannot be expected to name it accurately or describe how it is used, while patients with a primary naming disorder nevertheless will be able to tell you what an object is used for or otherwise demonstrate recognition. Typically, patients with naming disorders will tell you that they know what it is but cannot name it, while agnostic patients will be mystified as to what they are seeing or touching (depending on the type of agnosia) and will say they do not know what the object is. Patients with naming disorders usually are not helped by additional verbal descriptions of the object, while agnosics may have no difficulty naming an object from its verbal description. Mesulam (1985) describes the difference between an-

omias and agnosias as one of naming versus knowing. Agnosics do not know what an object is; anomics know but cannot name the object.

Agnosias are relatively rare disorders. In some cases, such as prosopagnosia (the isolated inability to recognize faces), the discovery of even one case justifies in-depth exploration of the patient’s abilities and publication of the findings. The simplest way to study the agnosias is to separate the agnosias by the sense or area of perception that is disrupted. This article will therefore explore visual, auditory, and somatosensory agnosias.

VISUAL AGNOSIAS

Visual agnosias are probably the most well studied, the first described, and the most easily recognizable of the agnosias. Munk, in the late 1800s, wrote of dogs that had no difficulty walking and avoiding obstacles but failed to react to things that had previously frightened or attracted them; this suggested that the dogs were able to sense objects but not recognize them (Bauer & Rubens, 1985). Although Kluver was initially unsuccessful in replicating this syndrome following occipital lobectomies in dogs, Kluver and Bucy in the 1930s were able to induce a similar syndrome in rhesus monkeys with bilateral temporal lobectomies (Kertesz, 1987). These monkeys placed virtually all objects in their mouths, including live snakes, which monkeys with intact recognition find extremely frightening.

Farah (1990) notes that this literature can be extremely confusing, presents examples of case studies in which the same patient is described by different authors as an example of mutually exclusive types of visual agnosia, and reviews some recent arguments that visual agnosia is not just rare but may not exist. Visual agnosias have been divided into two types since 1889, when Lissauer proposed a two-stage theory of recognition and classified agnosias as either apperceptive or associative; most researchers agree that this is a useful distinction (Bauer & Rubens, 1985).

Some researchers, such as Humphreys and Riddoch (1987), however, contend that this may be too elementary a classification system. Both apperceptive and associative agnosics have unimpaired elementary vision with intact acuity, brightness discrimination, and color

vision; but apperceptive agnosics, in addition to having impaired visual recognition, are unable to copy or identify even simple shapes.

Apperceptive Agnosias. Farah (1990) defines four types of apperceptive agnosia on the basis of behavior and lesion location. These are apperceptive agnosia in the most narrow sense, dorsal simultagnosia, ventral simultagnosia, and perceptual categorization deficits. Individuals who have apperceptive agnosia, narrowly defined, typically behave as if they were blind despite grossly normal visual fields, have roughly normal acuity, and are able to identify colors. They may maintain fixation and accurately determine which stimulus is closer than another (i.e., have intact depth perception), and yet be unable to copy, match, or recognize visual stimuli. Patients have some difficulty recognizing real objects, but much more difficulty recognizing line drawings and photographs. Identification of real objects seems to rely on the patient accurately guessing what an object is from cues based on color, texture, and size. Patients with apperceptive agnosia may spontaneously start tracing the outlines of stimuli and seem to benefit from this strategy in their ability to recognize simple geometric figures and written words. Others have been described who can identify a shape if it is drawn before them, apparently utilizing the cues from the movement, but are unable to identify the same figure if it is presented later. The neuropathology associated with this disorder is usually diffuse but typically involves the occipital lobe and surrounding areas. It has been demonstrated in patients following carbon monoxide poisoning, mercury poisoning, and head injury (Farah, 1990).

Simultagnosia was originally used to describe patients who seemed able to recognize individual elements of what they saw but were unable to incorporate them into a meaningful whole (Farah, 1990). Rizzo and Hurtig (1987) describe patients as "looking but not seeing," who complained of stationary objects that would disappear and of intermittent visual perception of their environment. Farah (1990) separates dorsal from ventral simultagnosia.

Patients with dorsal simultagnosia may correctly guess what they are seeing from extrapolating from the portion that they accurately perceive or may only be able to see one item in a scene with many details. One patient, to whom I showed a drawing of a typewriter,

appeared to see only one of the typewriter keys, which he thought resembled an olive. Initially guessing that the drawing was a martini, he quickly recanted, "Oh, but you wouldn't show me that." Other patients are described as spending many minutes examining drawings of various scenes, identifying each object individually without seeing the relations among the objects, or losing the perception of one object as they focus on another. These patients have extreme difficulty with reading or counting objects, having lost the ability to monitor what they have already read or counted. The size of the stimulus does not appear to be important, as Farah (1990) cites examples of patients unable to identify two small objects located very close to each other but able to read words regardless of the size of the printing. This suggests that the disorder relates to limitations in the patients' ability to attend to objects rather than just limitations in the size of the visual field. Patients with dorsal simultagnosia, like patients with narrowly defined apperceptive agnosia, tend to have bilateral, posterior lesions sparing the striate. Some cases of focal parieto-occipital lesions also have been shown to meet criteria for these disorders (Farah, 1990).

Ventral simultagnosia patients are similar to patients with dorsal simultagnosia in that they also recognize individual objects but do poorly when presented with multiple objects or more complex pictures, have difficulty with reading, and are not influenced by object size (i.e., they can see single objects regardless of size but cannot identify two small objects next to each other). Farah (1990) distinguishes between these two groups on the basis of the fact that ventral simultagnosia patients can see multiple objects even if they cannot recognize them. Individuals with ventral simultagnosia do not appear to be blind and can maneuver around objects in space and manipulate objects, guided by their vision. They have been referred to as "letter-by-letter" readers since they can read, albeit slowly, a letter at a time, and are able to count dots scattered across a page.

Patients with perceptual categorization deficit may not have clinically apparent problems, but the deficit can be diagnosed through experimental tasks. The hallmark of this disorder is an inability to recognize three-dimensional objects if the perspective is shifted. This inability was initially demonstrated in patients

with right-hemisphere lesions who could not match photographs of faces shot from different angles; later studies demonstrated that such patients were also unable to identify objects photographed from unusual perspectives (e.g., a ladder shot from below after having been laid on its side). Lesions in the right posterior inferior parietal lobe appear to be the critical ones for this disorder (Farah, 1990).

Associative Agnosias. Associative agnosia patients differ from those with apperceptive agnosia in their ability to produce complicated copies of drawings of high enough quality that others can clearly recognize the representation. Patients remain incapable of recognizing objects, however, either through naming them or through pantomiming their use. Apperception (perceptual analysis) is intact in these patients, as judged by their ability to identify the outlines of single items in overlapping line drawings, but they remain incapable of identifying the items (McCarthy & Warrington, 1986). These patients may describe individual features of a visually presented object and even recognize it as the same object when it is later presented from a different view, suggesting a stable visual representation of the object (Humphreys & Riddoch, 1987). They will quickly name and properly use these objects, once allowed to explore the object by touch, which argues against a primary language or naming problem or problem with knowing the use of objects. Little agreement is evident regarding the lesion site associated with this disorder. Farah (1990) notes that the majority of cases described have bilateral occipitotemporal damage, although various other authors have suggested that diffuse, nondiffuse, unilateral right, unilateral left, or bilateral lesions are necessary for this disorder. She attributes these discrepancies to different perceptual impairments, with different neuropathological lesions, all causing this same behavioral picture.

Farah (1990) classifies prosopagnosia, the inability to recognize faces, as a type of associative agnosia. The inability to recognize faces extends to both people whom the patient knew or recognized before the injury (e.g., family members, friends, and well-known celebrities) and people whom the patient meets after the injury. An extremely rare disorder, it is both devastating and fascinating. Patients identify a photograph as a face, describe the face and facial features (e.g., size of the nose or shape of the eyes), accurately say

whether the face is male or female, and may even identify facial expression and match the face with others showing similar expressions. L. F., described by Bauer (1982, 1984), once excused himself when he bumped into a full-length mirror, not recognizing the reflection as his own. The tragedy of this disorder was described by this same patient, a young man after he sustained his head injury in a motorcycle accident. He related that he was unable to make new friends, since people found it disconcerting and perhaps unbelievable that he appeared not to know them until they spoke, despite his clearly intact ability to see them. If he met a woman at a bar and she changed seats or put on a jacket over her clothes, he would be unable to identify her. To compensate, he, like many others described in this literature, attempted to use such cues as eyeglasses, hair color, or particular mannerisms to identify people. This strategy has limited utility, however, if everyone with bushy eyebrows is identified as Groucho Marx or if anyone with blue eyes seems to be one's blue-eyed mother.

The deficits associated with this disorder may not be limited to faces. Some patients are unable to distinguish between species of animals, and at least one farmer complained that he was no longer able to distinguish among his cows (Farah, 1990). Although initially it appeared that a bilateral lesion was necessary to develop prosopagnosia, Benton (1990) reviewed the literature and concluded that a right-hemisphere lesion can be sufficient to induce prosopagnosia, with a right inferior temporal lesion necessary but not sufficient to induce this disorder.

The inability to name only objects that are visually presented is called optic aphasia or visual object agnosia. The distinctiveness of this disorder also is controversial; some argue that it is a form of associative agnosia, and others insist that it is a naming disorder and not a true agnosia. What is distinct about this disorder is that patients can gesture and nonverbally indicate that they recognize an object but are unable to produce the name of the object, which may suggest a naming disorder; however, they are able to produce the correct name when provided with a definition or allowed to touch the object. Since some patients are able to name or describe the object after it is moved or rotated, this has also been labeled visual static agnosia (Mesulam, 1985). Farah (1990) localizes the le-

sion necessary for this disorder in the left posterior region, primarily the occipital cortex and white matter, while Mesulam (1985) contends that bilateral lesions in the occipitotemporal visual areas have been demonstrated most consistently in these patients.

Also called amnesic color blindness, color amnesia, and amnesia for color names, color agnosia is the inability to name colors, point to colors that an examiner names, or name the color of common objects (e.g., "What color is a fire truck?") despite intact perception of colors. Although some of these patients are able to identify incorrectly colored drawings, others are not (Lange, 1988). Mesulam (1985) disputes whether *color agnosia* is the most appropriate label for this "color-naming defect." Patients with this disorder should still describe the perception of color and have intact color matching. Lesions necessary for this disorder are described as mesial and in the left hemisphere between the temporal and occipital lobes (Mesulam, 1985).

AUDITORY AGNOSIAS

Acoustic agnosia, or auditory sound agnosia, refers to the inability to recognize sounds, despite intact perception and production of language. Wernicke's aphasia, in which both the ability to understand and the ability to produce language are impaired, is much more common. Both are believed to result from lesions of the posterior portion of the superior temporal gyrus. Just as sounds can be either verbal or nonverbal, auditory agnosias can be specific to verbal or nonverbal stimuli or affect both nonverbal and verbal recognition. Semantic associative agnosia, or nonverbal auditory agnosia, is the inability to recognize nonverbal sounds, such as sirens, ringing telephones, church bells, or car engines. Although both verbal and nonverbal auditory agnosia are associated with lesions to the temporal area, nonverbal auditory agnosia is most likely with right-hemisphere involvement (Mesulam, 1985). Not surprisingly, fewer patients are likely to present with complaints of this type of agnosia, as they may be unaware of their deficit, while most patients would quickly notice their inability to recognize spoken language.

Pure word deafness, auditory agnosia for speech, or auditory verbal agnosia is a specific deficit in the recognition of spoken language. Patients with this disorder

do not have a primary language disorder since they are able to read, write, and speak normally. Lip reading may help them understand spoken language, and they may be able to recognize who is speaking even when they are unable to understand what is being said. They may complain that what they hear is not clear or that it sounds like a foreign language. This syndrome has been hypothesized to result from the bilateral disconnection of Wernicke's area from auditory sensory input. Anatomically, most such patients have been found to have bilateral lesions of the anterior part of the superior temporal gyri. Some sparing, usually left sided, of Heschl's gyrus is most common (Bauer & Ruben, 1985).

Distinguishing between cortical auditory disorders and cortical deafness is extremely difficult, and some researchers have argued that this distinction is artificial. Both groups of patients are impaired in the recognition of all sounds (i.e., both verbal and nonverbal). Bauer and Rubens (1985) note that using the criteria that cortically deaf patients feel that they are deaf and behave as though they are deaf, while auditory agnosics are certain that they are not deaf, has not been particularly helpful. Patients with cortical deafness have consistently been found to have bilateral lesions of the auditory radiations or the primary auditory cortex, while more variability in lesion site has been found in auditory cortical disorder. Mesulam (1985) suggests that bilateral brainstem or diencephalic lesions may be responsible for cortical deafness, but notes that only a few cases have been described in the literature and more research is needed.

Assessment of one's perception of music is not part of the standard neurologic or neuropsychologic evaluation, but patients also have been described with specific deficits in their ability to recognize specific properties of music. Amusia is difficult to detect, since few patients complain of a specific loss of musical ability, and even more difficult to quantify given the great variability in musical abilities within the population (Bauer & Ruben, 1985). The degree of musical skill and training appears greatly to influence how musical abilities are organized in the brain. People with greater skill and training appear to utilize their dominant hemisphere more, which would suggest that left-hemisphere lesions would be more likely to disrupt musical abilities in these individuals. Left-hemisphere lesions

are associated with receptive amusia and difficulties with processing sequentially organized material, while expressive amusia is more common with lesions of the right hemisphere.

Patients with auditory affective agnosia are unimpaired in their ability to comprehend the content of speech, but are selectively impaired in their ability to comprehend the affective tone of spoken language. Heilman (1975) demonstrated that there are patients who are able to understand the content of spoken language but show impairment in the ability to recognize affect or the emotion conveyed by the nonverbal qualities of spoken language. He suggests that right temporoparietal lesions are responsible for this disorder, but Bauer and Rubens (1985) contend that additional research is needed before this disorder can be classified with confidence as true agnosia.

SOMATOSENSORY AGNOSIAS

Asterognosia, tactile agnosia, or tactile asymbolia is the inability to identify objects by touch, despite intact tactile sensation. First described by Wernicke in 1894, its existence has been disputed but is currently accepted (Goldberg, 1990). While some of these patients are unable to identify form and texture, others may retain the ability to distinguish between hard and soft; distinguish among cold, warm, and hot; and even identify an object as round and yet be unable to verbally or nonverbally indicate recognition of the object (Lange, 1988). Bauer and Ruben (1985) argue that this lesion is localized in the complex functional system in the middle third of the postcentral gyrus (the area governing the hand) and its cortical and subcortical connections.

Anosognosia is a disorder in the perception of the individual's own body and includes such things as neglect, denial of illness, phantom limb, and the inability to distinguish between one's own body and the body of the examiner. The parietal lobes—particularly the back of the right parietal area—seem to be implicated in these difficulties (Newcomb & Ratcliff, 1989). Patients with neglect typically do not attend to stimuli on the left side. These patients may neglect only tactile, visual, or auditory stimuli, or they may neglect any combination of senses. In some cases, the neglect is grossly apparent, with the patient consistently

bumping into tables and doorways on the neglected side. In other cases, the neglect is so subtle and mild that it is detected only when the patient is confronted with simultaneous bilateral stimulation and fails to notice ("neglects") the left-sided stimulation. Patients with anosognosia may tell you that they could move a paralyzed left limb if they wanted to or may deny that the affected limb is part of their body.

Newcomb and Ratcliff (1989) divide autotopagnosia (disorders of personal orientation) into three types: disorders of body part identification, finger agnosia, and deficits in right-left orientation. Disorders of body part identification, demonstrated by patients' inability to locate such things as the right eye or left ear, are associated with left anterior and left parieto-occipital lesions and are frequently associated with aphasia. Finger agnosia is the inability to use a finger independently or identify each finger separately, although the patient is able to use his or her fingers in such complex tasks as threading a needle. In Gerstmann's syndrome, patients show deficits in right-left orientation, writing, and calculations as well as finger agnosia, but there is considerable controversy regarding the "purity" of this syndrome and regarding whether other cognitive deficits may partially account for such a patient's difficulties. Patients with Gerstmann's syndrome may have lesions in either the left or the right hemisphere, but more regularly this disorder is associated with language comprehension problems. Deficits in right-left orientation, in addition to being found in Gerstmann's syndrome, have been described as a developmental disorder and an unusual symptom of parietal lobe dysfunction.

Sometimes included among the somatosensory agnosias are visual spatial agnosia, visual spatial dysgnosia, disorders of space perception, and unilateral spatial agnosia. Bauer and Rubens (1985), however, have argued that these conditions might be more appropriately classified as visuospatial disorders and hemispatial neglect.

CONCLUSION

Analysis of the ways cognitive abilities can be disrupted is invaluable in understanding not just brain dysfunction but also the mechanisms for normal perception. The study of the agnosias has been particu-

larly valuable in increasing our understanding of vision, hearing, and tactile sensations. The complexity of normal brain functioning is demonstrated by the fact that one single lesion or brain area is not implicated in all the agnosias and the finding that the same behavioral disturbances can be found after very different brain lesions. The literature on the agnosias can be confusing and at times contradictory; this highlights the need for additional study to resolve these disputes and make sense of these apparent contradictions. Additional research may assist in determining what abilities are necessary to perceive and recognize information in our environment and how perception can be disrupted after brain injury. As with other areas of neuropsychology, research so far in the agnosias raises as many questions about how the brain works as it answers.

(See also: PERCEPTION.)

BIBLIOGRAPHY

- BAUER, R. M. (1982). Visual hypoemotionality as a symptom of visual-limbic dysfunction in man. *Archives of Neurology*, 39, 702–708.
- BAUER, R. M. (1984). Autonomic recognition of names and faces in prosopagnosia: A neuropsychological application of the Guilty Knowledge Test. *Neuropsychologia*, 22, 457–469.
- BAUER, R. M., & RUBENS, A. B. (1985). Agnosia. In K. M. Heilman & E. Valenstein (Eds.), *Clinical neuropsychology*. New York: Oxford University Press.
- BENSON, D. F. (1989). Disorders of visual gnosis. In J. W. Brown (Ed.), *Neuropsychology of visual perception*. Hillsdale, NJ: Erlbaum.
- BENTON, A. (1990). Facial recognition in 1990. *Cortex*, 26, 491–499.
- CAMPION, J. (1987). Apperceptive agnosia: The specification and description of constructs. In G. W. Humphreys & A. J. Riddoch (Eds.), *Visual object processing: A cognitive approach*. Hillsdale, NJ: Erlbaum.
- DAMASIO, A. R. (1989). Reflections on visual recognition. In A. M. Galaburda (Ed.), *From readings to neurons*. Cambridge, MA: MIT Press.
- FARAH, M. J. (1990). *Visual agnosia: Disorders of object recognition and what they tell us about normal vision*. Cambridge, MA: MIT Press.
- FUNNEL, E. (1987). Objects and object names: Some deductions from acquired disorders of word processing. In G. W. Humphreys & A. J. Riddoch (Eds.), *Visual object processing: A cognitive approach*. Hillsdale, NJ: Erlbaum.
- GOLDBERG, E. (1990). Higher cortical functions in humans: The gradient approach. In E. Goldberg (Ed.), *Contemporary neuropsychology and the legacy of Luria*. Hillsdale, NJ: Erlbaum.
- HEILMAN, K. M., SCHOLES, R., & WATSON, R. T. (1975). Auditory affective agnosia: Disturbed comprehension of affective speech. *Journal of Neurology, Neurosurgery, and Psychiatry*, 38, 69–72.
- HUMPHREYS, G. W., & RIDDOCH, M. J. (1987). The fractionation of visual agnosia. In G. W. Humphreys & A. J. Riddoch (Eds.), *Visual object processing: A cognitive approach*. Hillsdale, NJ: Erlbaum.
- KERTESZ, A. (1987). The clinical spectrum and localization of visual agnosia. In G. W. Humphreys & A. J. Riddoch (Eds.), *Visual object processing: A cognitive approach*. Hillsdale, NJ: Erlbaum.
- KOLB, B., & WHISHAW, I. Q. (1985). *Fundamentals of Human Neuropsychology, Second Edition*. New York: W. H. Freeman.
- LANGE, J. (1988). Agnosia and apraxia. In J. W. Brown (Ed.), *Agnosia and apraxia: Selected papers of Liepmann, Lange, and Pottzl*. Hillsdale, NJ: Erlbaum.
- MARIN, O. S. (1987). Dementia and visual agnosia. In G. W. Humphreys & A. J. Riddoch (Eds.), *Visual object processing: A cognitive approach*. Hillsdale, NJ: Erlbaum.
- MCCARTHY, R. A., & WARRINGTON, E. K. (1986). Visual associative agnosia: A clinico-anatomical study of a single case. *Journal of Neurology, Neurosurgery, and Psychiatry*, 49, 1233–1240.
- MESULAM, M. (1985). *Principles of behavioral neurology*. Philadelphia: Davis.
- NEWCOMB, F., & RATCLIFF, G. (1989). Disorders of visuospatial analysis. *Handbook of Neuropsychology*, 2, 333–356.
- RIZZO, M., & HURTIG, R. (1987). Looking but not seeing: Attention, perception, and eye movements in simultagnosia. *Neurology*, 37, 1642–1647.

CAROL J. SCHRAMKE

ALCOHOL AND ALCOHOL ABUSE Alcohol consumption is widespread as a part of many religious, social, and recreational activities. It is esti-

mated that approximately 67 percent of the U.S. adult population drink alcoholic beverages. Fifty percent of the alcoholic beverages drunk are estimated to be consumed by only 10 percent of the population, however (U.S. Department of Health and Human Services, 1983). Thus, there appear to be two distinct groups of alcohol users: the majority of drinkers who use alcohol occasionally or infrequently, whom we might call social drinkers, and the minority of users, for whom alcohol use is a regular, frequent activity, often with significant negative consequences.

Clinically, the distinction between these two groups is made primarily on the basis of the physiological, psychological, and social consequences of alcohol use rather than on the quantity of alcohol consumed (American Psychiatric Association, 1987). Specifically, these consequences include legal problems, problems in the workplace, marital or relationship problems, and physical and mental health problems.

The term *alcoholic* may refer to an individual who is either a chronic abuser or physiologically/psychologically dependent on alcohol. National prevalence estimates obtained from Epidemiological Catchment Area data indicate that 10 to 15 percent of the adult population of the United States meet or have met the American Psychiatric Association standard diagnostic criterion for alcohol abuse or dependence (Regier et al., 1990).

This article will examine the acute and chronic effects of alcohol use on intelligence. Intelligent behavior is here defined as involving a variety of cognitive functions that may be independently affected by a variety of factors, including alcohol. When specific numbers of drinks are mentioned, a mixed drink is assumed to contain 1.5 ounces of 80-proof (40%) alcohol, and a 12-ounce beer is assumed to contain 4 percent alcohol; estimates regarding numbers of drinks assume a 150-pound person.

ACUTE EFFECTS OF ALCOHOL ON COGNITIVE FUNCTION

Ingestion of alcohol has been shown to affect immediately subsequent performance on various types of cognitive tasks differentially, depending on the specific task, the modality, and the timing and dose of the alcohol administration (Birnbaum et al., 1978; Goodwin

et al., 1969; Jones, 1973; Parker et al., 1980). In general, however, the acute effects of alcohol in moderate doses (e.g., 2.5 mixed drinks or 3 beers) to high doses (e.g., 5 mixed drinks or 6 beers) (Lister et al., 1991; Peterson et al., 1990) are reduction of response accuracy (Rundell & Williams, 1979), an adverse effect on decisionmaking (Mongrain et al., 1989; Peterson et al., 1990), impairment of abstracting performance (Lyvers & Maltzman, 1991), lengthening of reaction time (Huntly, 1973), and reduction of verbal fluency (Peterson et al., 1990). Some measures of cognitive function appear to be relatively immune to the acute effects of moderate doses of alcohol, including overall measures of the intelligence quotient (Peterson et al., 1990). In sum, it is clear that ingestion of as little as two beers can adversely affect many cognitive functions.

COGNITIVE PERFORMANCE IN SOBER SOCIAL DRINKERS

Does alcohol intake at the social-drinking level have residual effects in sober individuals? The research to date has been inconclusive. In several studies, moderate social drinkers (about eighteen beers or mixed drinks per month) have been shown to obtain lower scores than those of light drinkers (about three beers or mixed drinks per month) on some cognitive tests, but other researchers have not found these effects (Parsons, 1986a). Some investigations have also revealed neuroanatomical changes—specifically, some moderate social drinkers have shown brain shrinkage and perhaps differential retraction of dendrites (nerve endings) in certain areas of the brain (Harper & Kril, 1990). Since these data are highly variable and require replication, any conclusions must be considered tentative.

One reason for the discrepancies between studies is that different criteria are used to define "social" drinkers. Some studies have classified individuals as social drinkers if daily consumption levels fell below 80 grams of absolute alcohol (ethanol) per day (approximately four to five mixed drinks) (see Harper and Kril for examples). Others have used more conservative criteria, classifying individuals as social drinkers only if drinking levels fell between approximately 16 and 67 grams of ethanol per drinking day (one to three mixed

drinks) (Schaeffer & Parsons, 1986). Other reasons for the inconsistencies include potentially invalid reports of drinking levels and the differential effects of genetic and/or sex differences in responsivity to alcohol. Unfortunately, these inconsistencies in methodology make it difficult to predict the amount of regular alcohol consumption that will produce cognitive deficits in sober social drinkers.

COGNITIVE DEFICITS ASSOCIATED WITH THE CHRONIC ABUSE OF ALCOHOL

Organic Mental Syndromes. In its most severe form, alcoholism can result in organic mental syndromes: the Wernicke-Korsakoff Syndrome (WKS) and alcoholic dementia. Fortunately, these syndromes are relatively rare. It is estimated that only 10 percent of treated alcoholics meet the diagnostic criteria for either of these organic mental syndromes (Horvath, 1975). Although these syndromes constitute different nosological categories (disease classifications), they are similar in their symptomatology and are often difficult for the clinician to differentiate. In fact, a review of autopsy cases indicates that a high percentage of patients with alcoholic dementia also have diencephalic (midbrain) lesions, characteristic of WKS (Adams & Victor, 1989). WKS is produced by a thiamine deficiency accompanying chronic alcohol abuse in malnourished individuals (Victor, Adams, & Collins, 1971). Although a rather large percentage of alcoholics are malnourished, few alcoholics develop the disorder. This fact has led some researchers to propose a heritable susceptibility to thiamine deficiency, which may influence the progression and severity of the disorder (Martin, McCool, & Singleton, 1993).

WKS is defined by gross anterograde and retrograde memory loss. Patients with WKS also manifest minor perceptual impairment, deficits in problem solving and abstraction, and a loss of spontaneity and initiative (Victor, Adams, & Collins, 1971). They may be more impaired on measures of memory related to context-specific information ("episodic" memory) than on measures related to more general information ("semantic" memory), although both types of processes appear to be deficient (Butters & Cermack, 1980; Butters, Granholm, & Salmon, 1987).

Interestingly, these patients often have normal IQ levels (Talland, 1965). It has been hypothesized that the normal IQ levels are due to the relatively intact semantic memory function. WKS patients also exhibit relatively intact abilities to acquire procedural skills (e.g., learning motor tasks), although they are unable to recall having acquired the skill (Squire, 1987).

Intermediate-Stage Alcoholics. As noted, approximately 90 percent of sober alcoholics do not demonstrate the organic mental syndromes. These alcoholics have been referred to as intermediate-stage alcoholics. They frequently demonstrate significant alcohol-related cognitive deficits for some months and, in some cases, several years when compared to peer controls. Male and female alcoholics demonstrate similar patterns of cognitive deficits (Parsons, 1987a). This finding is particularly important given the shorter drinking histories reported by the majority of female alcoholics. These results, and others indicating earlier physiological dysfunction in female alcoholics, suggest that females may be more sensitive to the toxic effects of chronic alcohol abuse than males (Roman, 1988; Ross, 1989). Studies comparing IQ levels of chronic alcoholics with those of peer controls have indicated considerable variation. A review of the data, however, suggests a relatively consistent pattern—one that indicates that alcoholics' overall Wechsler Adult Intelligence Scale (WAIS) IQ levels are typically within normal range and may be only slightly lower than those of community controls. Considering the Verbal Scale and Performance Scale separately, however, suggests that although the Verbal IQ is relatively intact, the Performance IQ is frequently significantly lower (Parsons & Farr, 1981).

Further to discriminate the processes that are disrupted by chronic alcohol use, many studies have utilized tests that focus on specific cognitive functions. These studies have revealed alcohol-related cognitive deficits in a wide range of areas. Visual-spatial processing deficits have been obtained on tests such as the Block Design, Object Assembly, and the Picture Arrangement subtests of the Performance Scale of the Wechsler Adult Intelligence Scale-Revised (WAIS-R). These tasks require individuals physically to arrange objects to match the model, put together a jigsawlike puzzle, or place a series of pictures in logical order.

Tests of perceptual-motor skills have also revealed

poorer performance by alcoholics than by nonalcoholics. Tasks utilized in empirical studies have included the grooved pegboard, the Trail-Making task and Tactual Performance Test (TPT) from the Halstead-Reitan Battery (HRB), and the Digit Symbol (D-S) from the WAIS-R. The grooved pegboard requires subjects to insert grooved pegs in a board in a specific order. To complete the Trail-Making task, subjects trace a line through a series of alternating letters and numbers. The TPT requires subjects to complete a puzzle of blocks by feel. In completing the D-S, subjects draw in the symbol that has been assigned to each number as rapidly as possible.

Abstraction/problem-solving processes are frequently impaired in samples of detoxified alcoholics. Some of the standard tasks used to examine these processes include the Shipley Abstracting Subscale, the Wisconsin Card Sorting Test (WCST), and the category test from the HRB. These abstraction tasks require that subjects correctly identify the underlying "rule" or "concept" to complete the verbal items correctly (Shipley), correctly sort the cards (WCST), or correctly identify one of a series of four visuospatial figures according to some principle (Category test).

Although less consistently observed than deficits in the above area, alcohol-related deficits have also been observed on measures of learning and memory. A variety of tests have been used. Two of the more frequently used tests include the Wechsler Logical and Figural Memory scales. The Logical Memory Scale is a set of short paragraphs that subjects hear and immediately repeat verbatim, followed by a second delayed recall. The Figural Memory Scale is a set of line drawings that subjects draw immediately after viewing and then again at a delayed interval.

The tests listed above are only examples of the many tests administered within each of the functional areas. In general, it has been found that alcoholics manifest greater deficits on more difficult tasks for which performance is not dependent on highly overlearned information or processes. In Cattell's terms, alcoholics are more likely to exhibit deficits in fluid as opposed to crystallized intelligence (Cattell, 1963). A good review of tasks and functions is provided in Parsons (1987a).

Although the cognitive impairment exhibited by alcoholics relative to peer controls is usually statisti-

cally significant, these intermediate-stage alcoholics are rarely clinically impaired. More typically, the deficits are classified as mild to moderate in severity. Furthermore, 15 to 50 percent of these alcoholics fail to show significant cognitive deficits relative to the performance of age- and education-matched community controls (Parsons, 1986b).

Several neuropsychological hypotheses have been proposed to account for alcohol-related cognitive deficits. The right hemisphere hypothesis postulates that cognitive functions governed by the right hemisphere of the brain are impaired by alcohol use because this hemisphere is more sensitive to the toxic effects of alcohol than the left hemisphere. The frontal lobe-limbic system hypothesis postulates that the impaired cognitive functions are governed by the frontal lobe and limbic system and that this system is more sensitive to the toxic effects of alcohol than other regions of the brain. As of 1993, the mild generalized brain dysfunction hypothesis (Oscar-Berman, 1987) best accounted for the data. This hypothesis states that chronic alcohol abuse produces a mild to moderate global impairment, observed as a highly variable, non-specific pattern of cognitive dysfunction. Although this hypothesis can account for the data, like the other hypotheses, it lacks the specificity necessary to explanatory power and methodological rigor. As of 1993, models derived from work in cognitive science were beginning to be used (Glenn & Parsons, 1991; Nixon & Parsons, 1991). These models focus on underlying processes in addition to final performance.

Brain Changes in Alcoholics. A number of abnormalities in brain structure and function have been noted in chronic alcoholics. These abnormalities included brain shrinkage, a reduction in white matter, decreased cerebral blood flow, and changes in brain electrophysiology (Harper & Kril, 1990; Porjesz & Begleiter, 1987; Risberg & Berglund, 1987).

To some extent, these abnormalities appear to be reversible with continued abstinence. Age and gender may interact with this process, however: Younger alcoholics, female alcoholics, and alcoholics with shorter drinking histories exhibit signs of recovery from brain changes earlier than do other groups (Harper & Kril, 1990).

The correlation between neuropsychological and many of these neuroanatomical/neurophysiological

measures is quite modest. Perhaps as more sophisticated techniques become available, the nature of the relation between brain and cognitive function will be more clearly understood.

Recoverability of Cognitive Function. There is considerable recovery of function in chronic alcoholics with continued abstinence. The first cognitive skills recovered are those associated with verbal learning and memory, which are often recovered in the first month of abstinence. Research has suggested that other processes, such as those utilized in abstraction and problem solving, may remain inferior for as long as four years into sobriety (Parsons, 1987b). The most critical variable affecting recovery appears to be continued abstinence. Alcoholics who resume drinking, even if at levels lower than their pretreatment levels, exhibit less cognitive and neuroanatomical recovery (Parsons & Nixon, 1993).

CONCLUSION

Acute doses of alcohol have significant effects on cognitive processes—including perceptual-motor skills, visual-spatial processing, learning, memory, and problem solving and abstraction. The specific effects are contingent on the dose, the time of the administration relative to the learning and/or retrieval, and the type of cognitive process involved. There is no consistent research indicating cognitive deficits in sober social drinkers. Chronic use of alcohol in excessive amounts is related to long-lasting deficits in a wide variety of cognitive processes. WKS is evidenced in a small percentage of alcoholics who are susceptible to thiamine deficiency. It is marked by gross memory deficits, reduced spontaneity, and lack of initiative. Those alcoholics who do not meet criteria for WKS or other clinically diagnosable organic functional states often demonstrate significant deficits over the range of cognitive functions that are also impaired by acute doses of alcohol. There do not appear to be significant sex differences in the nature of these deficits, although some data suggest that females may be more susceptible to the negative effects of alcohol. A variety of brain changes accompany chronic alcoholism, but the correlations of these changes with cognitive deficits are low and variable. Some data suggest that recovery of

function is largely determined by continued abstinence: Alcoholics who resume drinking, even at reduced levels, do not demonstrate the same degree of cognitive recovery as do those who remain abstinent.

(See also: FETAL ALCOHOL SYNDROME.)

BIBLIOGRAPHY

- ADAMS, R. D., & VICTOR, M. (1989). *Principles of neurology* (4th ed.). New York: McGraw-Hill.
- AMERICAN PSYCHIATRIC ASSOCIATION (1987). *Diagnostic and Statistical Manual of Mental Disorders* (3rd Ed., Rev). Washington, DC: Author.
- BIRNBAUM, I. M., PARKER, E. S., HARTLEY, J. T., & NOBLE, E. P. (1978). Alcohol and memory: Retrieval processes. *Journal of Verbal Learning and Verbal Behavior*, 17, 325–335.
- BUTTERS, N., & CERMACK, L. S. (1980). *Alcoholic Korsakoff's Syndrome*. New York: Academic Press.
- BUTTERS, N., GRANHOLM, E., SALMON, D. P., ET AL. (1987). Episodic and semantic memory: A comparison of amnesic and demented patients. *JCEN*, 9, 479.
- CATTELL, R. B. (1963). Theory of fluid and crystallized intelligence: An initial experiment. *J. Educ. Psych.*, 105, 105–111.
- GLENN, S. W., & PARSONS, O. A. (1991). Effects of alcoholism and instructional conditions on speed/accuracy tradeoffs. *Alc: Clin Exp Res*, 15, 612–619.
- GOODWIN, D. W., POWELL, B., BREMER, D., HOINE, H., & STERN, J. (1969). Alcohol and recall: State-dependent effects in man. *Science*, 163, 1358–1360.
- HARPER, C. G., & KRIL, J. J. (1990). Neuropathology of alcoholism. *Alcohol & Alcoholism*, 25(2/3), 207–216.
- HORVATH, T. B. (1975). Clinical spectrum and epidemiological features of alcoholic dementia. In G. Rankin (Ed.), *Alcohol, drugs and brain damage* (p. 1). Toronto: Addiction Research Foundation.
- HUNTLY, M. S. (1973). Effects of alcohol and fixation task difficulty on choice reaction time to extrafoveal stimulation. *Journal of Studies on Alcohol*, 34, 89–103.
- JONES, B. M. (1973). Memory impairment on the ascending and descending limbs of the blood alcohol curve. *Journal of Abnormal Psychology*, 82, 24–32.
- LISTER, R. G., GORENSTEIN, C., RISHER-FLOWERS, D., WEINGARTNER, H. J., & ECKARDT, M. J. (1991). Dissociation of the acute effects of alcohol on implicit and

- explicit memory processes. *Neuropsychologia*, 29(12), 1205–1212.
- LYVERS, M. F., & MALTZMAN, I. (1991). Selective effects of alcohol on Wisconsin card sorting test performance. *British Journal of Addictions*, 86, 399–407.
- MARTIN, P. R., MCCOOL, B. A., & SINGLETON, C. K. (1993). Genetic sensitivity to thiamine deficiency and development of alcoholic organic brain disease. *Alc: Clin Exp Res.*, 17, 31–37.
- MONGRAIN, S., & STANDING, L. (1989). Impairment of cognition, risk-taking, and self-perception by alcohol. *Perceptual and Motor Skills*, 69, 199–210.
- NIXON, S. J., & PARSONS, O. A. (1991). Alcohol-related efficiency deficits using an ecologically valid test. *Alc: Clin Exp Res*, 15, 601–606.
- OSCAR-BERMAN, M. (1987). Neuropsychological consequences of alcohol abuse: Questions, hypotheses, and models. In O. A. Parsons, N. Butters, & P. E. Nathan (Eds.), *Neuropsychology of Alcoholism: Implications for Diagnosis and Treatment* (pp. 256–269). New York: Guilford Press.
- PARKER, E. S., BIRNBAUM, I. M., WEINGARTNER, H., HARTLEY, J. T., STILLMAN, R. C., & WYATT, R. J. (1980). Retrograde enhancement of human memory with alcohol. *Psychopharmacology*, 69, 219–222.
- PARSONS, O. A. (1986a). Cognitive functioning in sober social drinkers: A review and critique. *Journal of Studies on Alcohol*, 47(2), 101–114.
- PARSONS, O. A. (1986b). Alcoholics' neuropsychological impairment: Current findings and conclusions. *The Society of Behavioral Medicine*, 8(2–3), 13–19.
- PARSONS, O. A. (1987a). Do neuropsychological deficits predict alcoholics' treatment course and recovery? In O. A. Parsons, N. Butters, & P. E. Nathan (Eds.), *Neuropsychology of alcoholism: Implications for diagnosis and treatment* (pp. 273–290). New York: Guilford Press.
- PARSONS, O. A. (1987b). Neuropsychological consequences of alcohol abuse: Many questions—some answers. In O. A. Parsons, N. Butters, & P. E. Nathan (Eds.), *Neuropsychology of alcoholism: Implications for diagnosis and treatment* (pp. 153–175). New York: Guilford Press.
- PARSONS, O. A., & FARR, S. P. (1981). The neuropsychology of alcohol and drug use. In S. B. Filskov, & T. J. Boll (Eds.), *Handbook of Clinical Neuropsychology* (pp. 320–365). New York: Wiley.
- PARSONS, O. A., & NIXON, S. J. (1993). Neurobehavioral sequelae of alcoholism. *Neurologic Clinics*, 11, 205–218.
- PETERSON, J. B., ROTHFLEISCH, J., ZELAZO, P. D., & PIHL, R. O. (1990). Acute alcohol intoxication and cognitive functioning. *Journal of Studies on Alcohol*, 51(2), 114–122.
- PORJESZ, B., & BEGLEITER, H. (1987). Evoked pain potentials and alcoholism. In O. A. Parsons, N. Butters, & P. E. Nathan (Eds.), *Neuropsychology of alcoholism: Implications for diagnosis and treatment* (pp. 45–63). New York: Guilford Press.
- REGIER, D. A., FARMER, M. E., RAE, D. S., LOCKE, B. Z., KEITH, S. J., JUDD, L. L., & GOODWIN, F. K. (1990). Comorbidity of mental disorders with alcohol and other drug abuse. *Journal of American Medical Association*, 264(19), 2511–2518.
- RISBERG, J., & BERGLUND, M. (1987). Cerebral blood flow and metabolism in alcoholics. In O. A. Parsons, N. Butters, & P. E. Nathan (Eds.), *Neuropsychology of alcoholism: Implications for diagnosis and treatment* (pp. 64–75). New York: Guilford Press.
- ROMAN, P. M. (1988). Biological features of women's alcohol use: A review. *Public Health Reports*, 103, 628–637.
- ROSS, H. E. (1989). Alcohol and drug abuse in treated alcoholics: A comparison of men and women. *Alc: Clin Exp Res*, 13, 810–816.
- RUNDELL, O. H., & WILLIAMS, H. L. (1979). Alcohol and Speed-Accuracy Tradeoff. *Human Factors*, 21(4), 433–443.
- SCHAEFFER, K. W., & PARSONS, O. A. (1986). Drinking practices and neuropsychological test performance in sober male alcoholics and social drinkers. *Alcohol*, 3, 175–179.
- SQUIRE, L. (1987). *Memory and brain*. New York: Oxford University Press.
- TALLAND, G. (1965). *Deranged memory*. New York: Academic Press.
- U.S. DEPARTMENT OF HEALTH AND HUMAN SERVICES. (1983). *Fifth Special Report to the U.S. Congress on Alcohol and Health* (DHHS Pub. No. ADM 84–1291). Washington, DC: Superintendent of Documents, U.S. Government Printing Office.
- VICTOR, M., ADAMS, R. D., & COLLINS, G. H. (1971). *The Wernicke-Korsakoff Syndrome*. Philadelphia: Davis.

SARA JO NIXON
OSCAR A. PARSONS

ALZHEIMER'S DISEASE See BRAIN, PATHOLOGIES OF THE.

AMERICAN ASSOCIATION ON MENTAL RETARDATION The American Association on Mental Retardation (AAMR) is the oldest and largest professional organization devoted exclusively to issues related to mental retardation. The 9,000-plus members of this interdisciplinary association represent the United States and fifty-five other countries; they include professionals, as well as parents and consumers, whose common concern is the care, treatment, and education of persons with mental retardation.

Since its founding in 1876, the association's interests have included (1) institutionalization; (2) measures of intelligence and adaptation; (3) methods of training, educating, and caring for persons with mental retardation; (4) basic and applied research; (5) definitions, classification systems, and terminology; (6) informing and educating the public; and (7) advocating public policies on matters pertaining to mental retardation. The AAMR holds annual national conferences and publishes a monograph series and two scientific journals, *Mental Retardation* and the *American Journal on Mental Retardation (AJMR)*.

The organization was founded on June 6, 1876, in Media, Pennsylvania, by six physicians who represented institutions devoted to the care and education of "idiotic and feeble-minded children." A constitution was adopted naming the organization "The Association of Medical Officers of American Institutions for Idiotic and Feeble-minded Persons." Edouard Seguin, a prominent early leader in the field of mental retardation, was elected first president of the association. The name of the association was changed in 1906 to the "American Association for the Study of the Feeble-minded," in 1934 to the "American Association on Mental Deficiency," and in 1987 to the "American Association on Mental Retardation."

The relationship between the AAMR and the field of mental retardation has always been reciprocal. Both have been influenced by prevailing social, political, economic, and cultural factors. Consequently, positions of the AAMR have changed as the field has evolved. Nowhere is this reflected more than in changes in the descriptive terminology employed by the association over its history. Early association members used such terms as "idiot," "moron," "imbecile," "feeble-minded," "mental deficient," and "defective delinquent." Today the association advocates terminology

that emphasizes the individual rather than the disability; for example, "person with mental retardation" is preferred over "mentally retarded person."

Three areas that illustrate the interactive nature of the AAMR and society are (1) the role of institutions in the care and treatment of persons with mental retardation; (2) the influence of intelligence theory and testing on the field of mental retardation; and (3) the development of terminologies, definitions, and classification systems.

HISTORICAL BACKGROUND OF MAJOR ISSUES

The AAMR and Institutionalization. The primary interest of the association in its earliest years was the development and organization of institutions, or experimental schools, for training and educating "feeble-minded" children, who might then be returned to the community. Over time, members increasingly supported an expanded role for institutions. By the close of the nineteenth century, the overall consensus was that institutions should serve custodial as well as habilitative functions. A major purpose was to segregate all persons with mental retardation from society.

Prominent members of the association during the early 1900s included Henry Goddard, Walter Fernald, Frederick Kuhlmann, Lewis Terman, and Edgar A. Doll. Of these, perhaps none was more outspoken than Goddard who, convinced of the heritability of mental retardation along with concomitant devastating moral consequences for society, strongly advocated permanent sequestration of all persons with mental retardation, and compulsory sterilization of some, as a means of preventing hereditary transmission. This philosophy both influenced and reflected that of a society then in the midst of a period of "eugenic alarm." Consequently, during the first three decades of the twentieth century, new institutions were created and existing ones were enlarged at a rapid pace.

The AAMR and the Measurement of Intelligence. Paralleling this era in the association's history was the development of standardized tests to measure intelligence in school-aged children so that those who had difficulty learning could be identified early and receive specialized instruction. The idea originated in France, and in 1905 the task of developing

the measures fell to Alfred Binet, who invited his colleague, Théophile Simon, to collaborate. An American version of the "intelligence" test was developed by Lewis Terman of Stanford University in 1916 and became known as the Stanford-Binet Intelligence Test. In this revision, Terman used the term intelligence quotient (IQ) to represent the score achieved on the test (mental age divided by chronological age).

As the practice of administering group intelligence tests to schoolchildren and army recruits grew widespread in the United States, it became apparent that most individuals differed not in *kind* but in degree; that is, with the exception of those whose retardation derived from obvious organic pathology, there was not so much a dichotomous division between retarded and normal persons as there was a continuum of intellectual ability along which people differed in degree. In fact, there was a large group of persons who were neither "normal" nor mentally retarded, but "borderline." This discovery had a profound effect on the field of mental retardation and on the direction taken by the association.

First, it became apparent to the association that the population of "mental defectives" in the United States was far larger than previously believed. Efforts to sequester all in institutions would be futile. Thus, the association took a much more active interest in alternative placements and programs, including special classrooms within the regular school system, small, homelike halfway houses in the community, and parole for persons who had already been committed to institutions.

Second, the association was forced to examine its previous stand that being "mentally deficient" was synonymous with being "morally deficient." Before the introduction of intelligence tests, many outspoken members of the association had postulated a strong relation between "feeble-mindedness" and criminality, alcoholism, pauperism, and prostitution. When broad-scaled intelligence testing revealed large numbers of persons with mild retardation who had theretofore been unidentified, association members were struck by the realization that, contrary to popular wisdom, the majority of "mental defectives" lived within societal standards of right and wrong.

Third, the advent of intelligence testing crystallized the association's need to produce a comprehensive def-

inition of mental retardation and to classify levels of it. Previous definitions and classification systems had been based either on physical indicators or on the ability of the individual to adjust to social, educational, or other environmental demands. Now there was clearly a need to include level of intelligence as a descriptor.

The AAMR and Diagnosis, Classification, and Terminology. Almost from its inception, the association was aware of the need for a universally accepted definition of mental retardation. The major difficulty in devising such a definition was, then as now, due to the enormous heterogeneity among persons with mental retardation with respect to types, degrees, causes, and manifestations. The standardized intelligence test offered a quantifiable measure that could serve as a definitional underpinning. Despite the excitement and flurry of activity that the development of mental testing engendered, several prominent association members cautioned against exclusive reliance on the intelligence factor in the diagnosis of mental retardation.

In 1961, after decades of struggling with the problem of defining mental retardation, the association published a *Manual on Terminology and Classification in Mental Retardation*. The *Manual* contained the following definition:

Mental retardation refers to subaverage general intellectual functioning which originates in the developmental period and is associated with impairment in adaptive behavior.

Prior to the publication of the *Manual*, the association had supported a three-part classification scheme of mental retardation using the labels "idiot," "moron," and "imbecile" to describe, respectively, "low-grade," "middle-grade," and "high-grade" levels of retardation. The new *Manual* replaced this scheme with a five-part structure based on the following IQ ranges: Borderline (IQ 83–67), Mild (IQ 66–50), Moderate (IQ 49–33), Severe (IQ 32–16), and Profound (IQ < 16).

The addition of the "borderline" category aroused controversy among those within and outside the association. Opponents argued that if an individual's level of intellectual functioning was not clearly retarded, a diagnosis of mental retardation would be questionable. Proponents countered that the adaptive

behavior criterion, an essential feature of the new definition, would substantiate the “borderline” diagnosis. The intent of the criterion was to emphasize that an individual with low IQ was not to be judged mentally retarded if no significant impairment in adaptive behavior was demonstrated. A major problem with the criterion was the absence of valid measures.

Edgar A. Doll had, in 1953, developed and standardized the Vineland Social Maturity Scale, which measured the developmental behavior of youngsters, but no such scale existed for older individuals. In 1969, the association published the AAMD Adaptive Behavior Scale (ABS), an instrument normed on an institutionalized mentally retarded population. Because the original ABS was not generalizable to noninstitutionalized persons with mental retardation, a second version of the scale—applicable to an educational setting—was standardized on a sample of schoolchildren in grades second through sixth (or approximately 7–13 years old). The association published the second version in 1975. These two scales, together with the Vineland Social Maturity Scale, remain the most widely used measures of adaptive behavior.

THE AAMR FROM THE 1970s TO THE 1990s

The AAMR and Deinstitutionalization. The social upheaval that characterized the United States in the 1960s and 1970s directly influenced the modification of old positions and the assumption of new ones espoused by the association. Litigation brought by various minority groups, including advocates for persons with mental retardation, precipitated the new role of social and political activist for the association, which served as *amicus curiae* in several class action suits. Two important legal actions in the early 1970s that influenced the reversal of the association’s position on institutionalization were the landmark federal court case *Wyatt v. Stickney*, initiated in 1972, and the Education for All Handicapped Children Act of 1975 (Public Law 94-142).

The major issue in the *Wyatt* case was whether institutionalized persons with mental retardation had a constitutional “right to treatment.” The argument was that civil commitment of persons with mental retar-

ation to institutions was equivalent to penal incarceration if habilitative treatment was not accomplished. The complex litigation spanned a decade and produced a number of decisions regarding the rights of persons with mental retardation, among them the right to receive appropriate treatment in the “least restrictive setting.”

The Education for All Handicapped Children Act invoked the principle of placement of handicapped children in the “least restrictive educational setting.” “Least restrictive,” in both cases, referred to settings that were most like those available for persons without mental retardation if the most appropriate treatment for the individual could be provided in such a setting.

Neither *Wyatt* nor the passage of Public Law 94-142 was instigated by the AAMR. However, the association quickly embraced the concept of “least restrictive environment” and, in one of the most salient illustrations of a reversal in position based on the prevailing social climate, became an active advocate of the shift from institutional to community programs for the mentally retarded. The association has maintained an interest in existing institutions and has played a significant role in the development of standards for residential facilities. Nevertheless, one of the original objectives of the association, “[to] lend its influence to the establishment and fostering of institutions . . .,” is missing from more recent mission statements, which include “promoting high quality services and supports that enable full community inclusion [of persons with mental retardation].”

Definitions and Classification Systems. The 1961 *Manual on Terminology and Classification in Mental Retardation* was updated and revised in 1973, 1977, 1983, and 1992. The “borderline” classification was dropped in the 1973 revision. The upper limit of intellectual functioning still considered to be within the range of retardation was an IQ of about 70 as determined from a standardized general intelligence instrument. No substantial change was made in the definition or classification scheme after 1973 until the 1992 revision. Both the title of the 1992 version, *Mental Retardation: Definition, Classification, and Systems of Supports*, and the definition it contains reflect the evolving philosophy of the association. The definition states that

Mental retardation refers to substantial limitations in present functioning. It is characterized by significantly subaverage intellectual functioning, existing concurrently with related limitations in two or more of the following applicable adaptive skill areas: communication, self-care, home living, social skills, community use, self-direction, health and safety, functional academics, leisure, and work. Mental retardation manifests before age 18.

Although the initial diagnosis of mental retardation still relies on measured intelligence, the previous subcategories of "mild," "moderate," "severe," and "profound" levels of retardation are no longer used. The 1992 revision emphasizes instead interactions with the environment in which individuals must function as well as the support needs they might require for improved functioning. Because the definition and classification system advanced by the AAMR has traditionally been the most widely employed in the United States, the change in focus from level of disability to supports, abilities, natural environments, and empowersments has broad implications for professionals in the field as well as for consumers.

CONCLUSION

From an original membership of six physicians interested in the care and treatment of persons with mental retardation, the AAMR had by the early 1990s evolved into an international organization of over 9,000 members from a wide variety of disciplines. Active participation on the part of members has been encouraged by the creation of regional, state, and local chapters.

Positions of the association have often been influenced by prevailing social and cultural conditions. Perhaps the most pessimistic era in the association's history was the period from 1900 to 1930, when "mental deficiency" was thought to be strongly linked to criminality and other social ills. During this time, policies such as compulsory sterilization, sequestration, and restriction of marriage were promoted to curtail hereditary transmission of "feeble-mindedness." These views lingered until the post-World War II period and served to perpetuate negative stereotypes. Since that time the AAMR has worked to dispel the stereotypes

and to help create a supportive communal environment for persons with mental retardation.

Many individual members of the association have contributed to an understanding of intelligence, though the role of the association itself has been to use, rather than produce, theory and knowledge pertaining to intellectual functioning. The development of intelligence tests, for example, provided the foundation for a universally accepted definition of mental retardation. Though the association never formally challenged the theoretical basis of "general" intelligence as reflected in an IQ score, it did recognize the need to view mental retardation in the context of adaptation. Thus, the concept of adaptive behavior, included in the definition of mental retardation, requires that cultural relevance and environmental factors be considered integral elements of the diagnostic process along with intellectual functioning.

The movement toward deinstitutionalization and the right to treatment in the least restrictive setting necessitated the development of alternative programs and placements for persons with mental retardation. This, in turn, required an assessment of the interaction between persons who have mental retardation and the environment in which they function. The 1992 AAMR position reflected these changes by categorizing the supports needed to facilitate the functioning of persons with mental retardation rather than by categorizing the people themselves.

Research interests, as reflected in the journals of the association, include interpretations of intelligence test scores, the relation between learning and IQ, analyses of behavior excesses, social factors, and early intervention. Attention also has been paid to factors that predispose mental retardation such as low birth weight and toxin exposure.

BIBLIOGRAPHY

- BAUMEISTER, A. A., & BUTTERFIELD, E. (1970). *Residential facilities for the mentally retarded*. Chicago: Aldine.
- GROSSMAN, H. J. (ED.). (1983). *Classification in mental retardation*. Washington, DC: American Association on Mental Deficiency.
- HEBER, R. (1961). *A manual on terminology and classifica-*

tion in mental retardation (2nd ed.). Monograph supplement to the *American Journal of Mental Deficiency*.

LUCKASSON, R. A., COULTER, D. L., POLLOWAY, E. A., REISS, S., SCHALOCK, R. L., SNELL, M. E., SPITALNIK, D. M., & STARK, J. A. (1992). *Mental retardation: Definition, classification, and systems of support*. Washington, DC: American Association on Mental Retardation.

SCHEERENBERGER, R. C. (1987). *A history of mental retardation*. Baltimore: Paul H. Brookes.

SLOAN, W., & STEVENS, H. A. (1976). *A century of concern: A history of the American Association on Mental Deficiency 1876–1976*. Washington, DC: American Association on Mental Deficiency.

PAMELA WOODLEY-ZANTHOS
ALFRED A. BAUMEISTER

AMERICAN COLLEGE TEST (ACT) The original ACT Assessment was first administered nationally in the fall of 1959. It was designed to measure the broad academic skills developed through formal secondary school coursework in English, mathematics, social studies, and natural science. Scores were reported on a scale of 1 to 36 for English usage, mathematics, social studies reading, and natural sciences reading; a composite score, the arithmetic average of the four scaled scores rounded to the nearest whole number, was also reported.

By 1988, more than a million high school juniors and seniors were being tested annually on five national test dates. ACT scores were being used for admissions, academic advising, career exploration, and course placement in over 90 percent of U.S. colleges and universities.

In 1989, an enhanced edition of the ACT Assessment was introduced, containing four curriculum-based tests that measure academic achievement in the areas of English, mathematics, reading, and science reasoning. The tests are based on and oriented toward the major areas of secondary and postsecondary instructional programs. Performance on these tests has a direct and obvious relationship to a student's academic development. Furthermore, the meaning of that performance, or score, can be readily grasped and interpreted by both instructional staff and students.

The fundamental idea underlying the development and use of these four tests is that the best way to measure a student's preparedness to benefit from postsecondary education is to measure as directly as possible the knowledge and skills students will need in that setting. The specific knowledge and skills selected for evaluation were determined through a detailed analysis of three sources of information. First, the objectives for instruction for grades seven through twelve were obtained for all states that had published such objectives. Second, textbooks on state-approved lists for courses in grades seven through twelve were reviewed. Third, educators at the secondary and postsecondary levels were consulted to determine which knowledge and skills taught in grades seven through twelve were required for successful performance in postsecondary courses. These three sources of information were analyzed to define a scope and sequence for each of the areas measured by the ACT Assessment.

The specifications for the ACT Assessment tests were developed from this achievement continuum with the assistance of nationally recognized educational consultants. The goal was to select for assessment knowledge and skills taught at the secondary level that are important for success in postsecondary education.

The tasks presented in the tests are representative of a broad range of academic skills, comprehensive in scope, and educationally significant. The four tests are measures of academic development that rely largely on the student's skill in applying to higher-level tasks the content knowledge and reasoning skills acquired through coursework. The tasks often require the integration of proficiencies and skills from various high school courses. Consequently, the ACT Assessment tests contain a large proportion of analytical problem-solving exercises and relatively few measures of narrow skills or basic recall.

The materials for the ACT Assessment tests are produced by item writers who represent a wide variety of backgrounds. In the construction of its tests, ACT conscientiously involves educators of both genders from educational institutions in all regions of the country, who reflect a variety of racial and ethnic backgrounds and represent different educational phi-

losophies. These item writers work from detailed guidelines that specify the test content, cognitive skill level, item format, and fairness criteria used to construct the ACT Assessment tests.

ACT SUBJECT AREA TESTS

The ACT Assessment includes four tests:

- English 75 items, 45 minutes
- Mathematics 60 items, 60 minutes
- Reading 40 items, 35 minutes
- Science Reasoning 40 items, 35 minutes

(Total testing time: 2 hours, 55 minutes)

English. This seventy-five-item test measures the student's understanding and use of written English. The format consists of five prose passages, with a series of test questions related to each passage. The passages are designed to assess six elements of effective writing in two broad areas, usage/mechanics (punctuation, basic grammar and usage, sentence structure), and rhetorical skills (strategy, organization, style). The scale score has a range of 1 to 36; subscores are reported for usage/mechanics and rhetorical skills, on a scale of 1 to 18.

Mathematics. The sixty-item mathematics test is designed to assess students' skills at solving practical mathematics problems. The test content utilizes skills gained in pre-algebra, elementary algebra, intermediate algebra, coordinate geometry, plane geometry, and trigonometry. The mathematics scale score has a range of 1 to 36; subscores are reported for pre-algebra/elementary algebra, intermediate algebra/coordinate geometry, and plane geometry/trigonometry, on a scale of 1 to 18.

Reading. This forty-item test is designed to assess the student's level of reading comprehension. Each of four prose passages represents the type of text found in college freshman courses. The scale score has a range of 1 to 36; subscores are reported for social studies/sciences and arts/literature reading skills, on a scale of 1 to 18.

Science Reasoning. This forty-item test measures the student's ability to interpret, analyze, evaluate, reason, and solve problems in the natural sciences. Test items are based on information from three types

of format: data representation, research summaries, and conflicting viewpoints. The scale score has a range of 1 to 36; there are no subscores.

SCORES

Students receive twelve scores on the ACT Assessment—four test scores, seven subscores, and a composite score:

English Scores

Subscores

Usage/Mechanics

Rhetorical Skills

Mathematics Score

Subscores

Pre-Algebra/Elementary Algebra

Intermediate Algebra/Coordinate Geometry

Plane Geometry/Trigonometry

Reading Score

Subscores

Social Studies/Sciences

Arts/Literature

Science Reasoning Score

Composite Score

The composite score is the arithmetic average of the four test scores, rounded to the nearest whole number.

OTHER PARTS OF THE ACT ASSESSMENT

Student Profile. This questionnaire collects a wide range of information, including:

- admissions/enrollment information
- educational plans, interests, and needs
- special educational needs, interests, and goals
- college extracurricular plans
- need for financial aid; work plans
- background information
- factors influencing college choice
- high school information
- high school extracurricular activities
- out-of-class accomplishments
- evaluation of high school experience

Interest Inventory. Because many students explore career options as they plan for college, a system-

atic assessment of career interests has been part of the ACT Assessment for more than fifteen years. Items in the current inventory have been updated through a series of field studies involving 12,000 ninth-graders, college-bound students, and adults. The Interest Inventory continues to use ACT's World-of-Work Map and Career Family List to link students' interest scores to educational and career options.

High School Courses and Grades. When students register for the ACT Assessment, they report the last grades they earned in thirty specific college-preparatory courses. This self-reported information about coursework is used to determine whether a student has or will have completed "core" or "less than core" coursework. Core coursework is defined as four or more years of English, three or more years of mathematics beyond pre-algebra, three or more years of social studies, and three or more years of science.

NORMS, SERVICES, AND TECHNICAL INFORMATION

ACT publishes yearly updated norms based on all ACT-tested high school graduates. College-bound norms based on a 1988 national norming study are also available. Currently, about 35 percent of U.S. high school seniors take the ACT Assessment. The average ACT composite score for 1992/1993 ACT-tested high school graduates was 20.7.

Because the tests measure knowledge and skills deemed necessary for success in college, students who take a more rigorous high school program tend to earn higher ACT scores than do those who do not. This pattern of achievement according to type of high school program is consistent across ethnic group, gender, and ethnic group within family income levels. Students who take a core program also tend to be more successful during the first year of college, as measured by both ACT scores and college grades. Summary profile information on ACT-tested students is provided annually to participating high schools, colleges, universities, and state agencies. Course placement and recruitment services are available to colleges and universities.

Because the ACT Assessment is a secure examination, a number of forms of the tests are used each year. The reliability of the composite score for each form is

about .96. Further details about the technical aspects of the AAP are contained in the documents listed in the bibliography.

USES OF THE ACT ASSESSMENT

Since 1959, the ACT Assessment Program has been an important part of the high school to college transition. Each year, nearly 1.5 million students take the ACT Assessment to help them identify and develop realistic plans for accomplishing their educational and career goals as they move from secondary to postsecondary education. Students use ACT Assessment information in planning for college and in presenting themselves to colleges as persons with unique patterns of educational development, accomplishments, and needs.

High school counselors use ACT information in counseling students about their probability of success at specific colleges, and in exploring with them possible majors and career choices. Colleges use ACT data in recruitment and in making decisions about admissions, course placement decisions, and scholarships. College advisers use ACT information to help students select majors and courses.

All but about thirty U.S. postsecondary institutions require, recommend, or accept ACT scores. Concordance studies have been conducted over the years to determine the association of ACT scores with scores from the SCHOLASTIC ASSESSMENT TESTS (SAT), which has different specifications. Typically, the correlation between the ACT composite score and SAT total (e.g., verbal plus quantitative) is about .89. A reference for the most recent comprehensive study of concordance between ACT and SAT scores appears in the bibliography.

BIBLIOGRAPHY

- ACT Assessment User Handbook* (Updated yearly). American College Testing Program, Iowa City, IA.
- EXPLORE Interpretive Guide for Student and School Reports*. (1992/1993). American College Testing Program, Iowa City, IA.
- HOUSTON, W., & SAWYER, R. (1991). Relating scores on the Enhanced ACT Assessment and the SAT Test Batteries. *College and University*, 66(4), 195-200.

MARCO, G., & OBEDEL-FATTAH, A. A. (1991). Developing concordance tables for scores on the Enhanced ACT Assessment and the SAT. *College and University*, 66(4), 187–194.

PLAN Technical Manual. (December 1989). American College Testing Program, Iowa City, IA.

Preliminary Technical Manual for the Enhanced ACT Assessment. (October 1989). American College Testing Program, Iowa City, IA.

Supplement to the Preliminary Technical Manual for the Enhanced ACT Assessment. (December 1991). American College Testing Program, Iowa City, IA.

JAMES MAXEY

ANALYTICAL ABILITY See REASONING, DEDUCTIVE; REASONING, INDUCTIVE.

ANASTASI, ANNE (1908–) Anne Anastasi was born in 1908 in New York City to Anthony and Theresa Gaudiosi Anastasi. Her father, who worked for the New York City Board of Education, died when she was 1 year old. Anne was raised by her mother, her maternal grandmother, and her uncle.

Anne Anastasi was educated at home under the tutelage of her grandmother until the age of 9. Following elementary school, Anastasi was enrolled for two years in the Rhodes Preparatory School in Manhattan. She was admitted to Barnard College in 1924 at the age of 15, and graduated four years later with a bachelor's degree in psychology. She received the Caroline Duror Graduate Fellowship, awarded "to that member of the graduating class who shows greatest promise of distinction in her chosen line of work" (Sexton & Hogan, 1990, p. 15). She received her doctoral degree in experimental psychology in 1929 at the age of 21 from Columbia University, where she studied under Harry Hollingworth, Robert Woodworth, Clark Hull, and H. E. Garrett, who was her dissertation mentor. During her doctoral training, she served as research assistant to Charles B. Davenport, working on the development of culture-free tests. She was influenced by the work of Otto Klineberg on racial differences and intelligence testing. An article by Charles SPEAR-

MAN about correlation coefficients convinced her to combine her interests in mathematics and psychology.

In 1930, Anastasi was made an instructor of psychology at Barnard. She stayed at Barnard until 1939 when she became assistant professor and chair of the new department of psychology at Queens College of the City University of New York. She remained at Queens College until 1946. In 1947 Anastasi was appointed associate professor of psychology at Fordham University where she ultimately attained the rank of professor. She remained at Fordham until her retirement in 1979. She was then given professor emeritus status and was awarded an honorary Doctor of Science degree. Additional honorary doctorates were bestowed on her from Cedar Crest College, LaSalle College, Villanova University, and the University of Windsor in Canada.

Anastasi's accomplishments have earned her a worldwide reputation as a prominent psychologist and researcher. She is considered by many to be one of the leading psychologists in the English-speaking world, having devoted her life to this discipline. She is a researcher, a teacher, an organizational leader, and a textbook writer. Her books are used extensively throughout the world. Her most famous text, *Psychological Testing*, is so popular and is so concisely written and understandable that it alone would have given her acclaim. This text was first published in 1954 and is now in its sixth edition. She also wrote other major texts, including *Differential Psychology* and *Fields of Applied Psychology*. These three works were cited as being "models of clarity, comprehensiveness, and synthesis" when she won the 1984 American Psychological Foundation Gold Medal Award for lifetime achievement. Anastasi also edited *Testing Problems in Perspective*, a collection of papers presented over a twenty-five-year span at the Annual Invitational Conference on Testing Problems, sponsored by Educational Testing Service. In addition to her textbook writing and editing, Anastasi has contributed many reviews to the *Mental Measurements Yearbook (MMY)*. With the publication of the eighth *MMY* yearbook in 1978, Anastasi became the only contributor to have written reviews in all *MMY* volumes to that date from its inception in 1938, and she considered it her "privilege to have been among the many psychologists participating in this monumental project" (Anastasi, 1972, p. 12).

As a researcher, Anastasi has investigated a variety of themes. Her primary work has focused on intelligence and ability testing, its evaluation and interpretation. She has investigated the misconceptions about "culture free" or "culture fair" assessment. She has written and spoken extensively on the methodological problems of test bias, speeded tests, coaching, and intelligence-test item selection. As such, Anastasi has established herself as an authority on the subject of psychological testing. This reputation led to her invitation to serve as an individual consultant for the College Entrance Examination Board, the American College Testing Service, and the Educational Testing Service.

Anastasi had a long-standing interest in the study of individual differences and the formation of trait characteristics, beginning with her dissertation in 1930. To this end, she advanced the development of psychology as a quantitative behavioral science. Her investigation of language development in black and Puerto Rican children considered the role of experience. She also incorporated experiential factors in investigations of family size and intelligence, age changes in adult test performance, and gender differences in psychological traits. The role of experience was also examined in her investigation of creativity in children and adolescents.

In 1972, Anastasi was the third woman to be elected as president of the American Psychological Association. She served as president of the Eastern Psychological Association (1946–1947), as well as president of the APA Divisions of General Psychology (1956–1957), and Evaluation and Measurement (1965–1966). In 1984 Anastasi was honored by the American Psychological Foundation. She was granted that year's Gold Medal Award in recognition of her long and continued record of scientific and scholarly accomplishments. In 1987, she was presented the National Medal of Science by President Ronald Reagan.

Much of what we currently know concerning psychological traits, the construction and interpretation of psychological tests, and the experiential and environmental influences on development have been directly influenced by the writing and research of Anne Anastasi. Her long and impressive record of accomplishments demonstrates relentless diligence and dedication.

BIBLIOGRAPHY

- AMERICAN PSYCHOLOGICAL ASSOCIATION. (1982). Distinguished scientific award for the applications of psychology: 1981. *American Psychologist*, 37, 52–59.
- AMERICAN PSYCHOLOGICAL ASSOCIATION. (1985). American Psychological Foundation Awards for 1984. *American Psychologist*, 40, 340–345.
- ANASTASI, A. (1972). Reminiscences of a differential psychologist. In T. S. Krawiec (Ed.), *The Psychologists* (Vol. 1). New York: Oxford University Press.
- ANASTASI, A. (1980). Anne Anastasi. In G. Lindzey (Ed.), *A history of psychology in autobiography* (Vol. VII). San Francisco: W. H. Freeman.
- ANASTASI, A. (1988). Anne Anastasi, 1908–. In A. N. O'Connell & N. F. Russo (Eds.), *Models of achievement: Reflections of eminent women in psychology* (Vol 2). Hillsdale, NJ: Erlbaum.
- DENMARK, F. L., & FERNANDEZ, L. C. (1992). Women: Their influence and their impact on the teaching of psychology. In A. E. Puente, J. R. Matthews, and C. L. Brewer (Eds.), *Teaching psychology in America: A History*. Washington, DC: American Psychological Association.
- SEXTON, V. S., & HOGAN, J. D. (1990). Anne Anastasi. In A. N. O'Connell & N. F. Russo (Eds.), *Women in psychology: A bio-bibliographic sourcebook* (pp. 13–22). New York: Greenwood Press.

FLORENCE DENMARK

ANIMAL INTELLIGENCE: HISTORICAL PERSPECTIVES AND CONTEMPORARY APPROACHES

Interest in animal intelligence has a long history, traceable to at least the ancient Greek civilizations (Dewsbury, 1984). This history is largely a struggle between discrepant views of the relations between humans and nature, and animals and humans. Many Greek philosophers espoused a continuity within all nature. For example, Anaxagoras believed that all animals have intelligence, but he underscored the superiority of human intellect. Similarly, Aristotle arranged the animal species in a hierarchy (*Scala naturae*) according to their complexity. Man, with the most complex development of mental traits, was at the top.

During the Middle Ages (sixth to sixteenth centuries C.E.), Christian theology dominated thinking in Europe. The doctrine of special creation replaced the view of continuity in nature. This doctrine held that humans were a special product of divine creation, not part of nature, and therefore inappropriate for scientific study. Furthermore, the doctrine inexorably separated animals and humans by the presence of a soul in humans and by the human's capability for reason.

In the seventeenth century, René Descartes championed the doctrine of "dualism," that the universe consisted of two independent, noninteracting entities: mind and matter or, for the human, mind and body. As an extension of this view, he separated human behavior into actions that were voluntary (a product of the mind) and actions that were involuntary (a product of "animal spirits" acting strictly within the domain of the body, and mechanical or reflex-oriented in nature). Descartes upheld many of the views from the Middle Ages concerning what distinguished humans from animals, including the idea that all "intelligent" behavior was voluntary, a product of the soul found only in humans. Because involuntary behavior had nothing to do with the soul, however, it could be studied equally well in humans or animals.

In the late nineteenth century, Charles Darwin proposed the theory of the evolution of all living things, including humans. The keys to the theory were the ideas that heritable traits are subject to genetic variation; that those species whose members possess variations most adaptive to their environmental niche will survive and proliferate, passing on these variations to other generations; that humans are not products of special creation, but rather are a part of nature and subject to the same evolutionary laws as other species; that the mind and body are not, as Descartes had proposed, completely separate from each other; and that continuities with human traits can be found in other animals. As a logical extension of his findings on phylogenetic and structural continuities, Darwin proposed that behaviors, mental abilities, and mental states formerly attributed to humans only also existed in the animal kingdom, in species that either shared a common ancestry with humans or that had been subject to similar environmental and social pressures. Darwin stated that the difference between the mind of a hu-

man and that of the highest animal was one of degree and not of kind. This consequently opened all behavior of humans and animals to scientific investigation, including mental traits and skills.

Darwin's theories aided in the development of the field of comparative psychology—the scientific investigation of the behaviors and mental abilities of different animals. His views inspired others to search for mental continuities in such advanced cognitive abilities as reasoning, abstraction, and problem solving, and to attempt to trace the development of these abilities in different animal species. Some proponents of Darwin's theory postulated continuities in a wide variety of mental traits among most vertebrates. These "vitalists," as they were termed, believed in the existence of a "vital" force peculiar to living organisms. They attempted to reveal the mental continuity of humans and animals by collecting large bodies of anecdotal evidence. George Romanes's book *Animal Intelligence* (1882), for example, is replete with anecdotes of behaviors of animals that he attributed to conscious thinking. Vitalists also used the method of introspection to reveal mental events in others (i.e., determining another's mental state by observing their behavior and inferring one's own mental state when engaging in similar behavior).

Because the methods employed by the vitalists relied on introspection, their views on the broad continuity of mental characteristics came under criticism from a more conservative group of scientists and philosophers—the "mechanists." Mechanists adopted a Cartesian attitude, preferring to view most animals as automatons driven by instinct and reflex. Jacques Loeb contended that mental traits such as consciousness, thinking, or abstraction were beyond the capabilities of nearly all animals. Furthermore, Loeb contended that much of animal behavior, which on the surface appears to be the result of conscious thought, is in reality a product of simple reflex actions emitted in response to environmental stimuli.

Partially as a reaction against the vitalist approach to the study of "mind" in animals and the broad inferences deriving from this approach, C. Lloyd Morgan wrote in his 1894 book *An Introduction to Comparative Psychology* that "in no case may we interpret an outcome as the exercise of a higher psychical faculty, if it

can be interpreted as the outcome of the exercise of one which stands lower in the psychological scale.” This caveat, referred to in later years as “Morgan’s Canon,” provided an important “check” on the interpretation of observed animal behavior that on the surface may appear to be the result of higher intellectual mechanisms but can be attributed also to simpler elementary processes.

Edward THORNDIKE, in his book *Animal Intelligence* (1898), also railed against subjective methods for interpreting animal behavior. Thorndike proceeded to demonstrate through careful experimentation the way in which animals learn through trial and error and the influence of the outcomes of a response on the probability of repeating that response. He proposed that those responses closely followed by satisfaction to the animal would be more likely to recur and that those followed by discomfort would be less likely to recur. This “law of effect” set the stage for a subsequent emphasis on experiments on learning processes and learning theory that was to continue well into the twentieth century.

From the early to mid-twentieth century, the study of mental processes and mental events in animals fell into disrepute among most American psychologists. In 1913 John Watson published a landmark paper called “Psychology as the Behaviorist Views It.” In this paper and in later writings Watson championed the view that although mental states and mental processes may be present in animals, they are unobservable. He urged that the scientist should study only observable events, and for the psychologist, this meant observable behavior. Watson’s views, as well as Thorndike’s methods, inspired a large following among psychologists interested in animal behavior. Mental terms, such as *memory* or *thinking*, became unacceptable within the behavioristic literature. Psychologists instead sought to uncover the universal laws of learning, which they believed characterized all animal species. One of the best known and most influential proponents of this “behaviorist” approach was B. F. Skinner. Skinner was largely responsible for the development of what has been termed *radical behaviorism*. Some of the tenets of radical behaviorism are:

1. The appropriate task of psychology should be the study of behavior.

2. The purpose of such study should be to discover lawful relations or associations between environmental stimuli (S) and resultant behavioral responses in an organism (R).
3. All behaviors can be described in S–R terms.
4. Reinforcers control the formation of S–R associations and are universal in their applicability.
5. Given an adequate environmental description and the lawful relations between environmental variables and behavior, all behaviors are predictable.
6. All behaviors, no matter how complex, follow the same principles of learning regardless of the species under study or the task being conducted.
7. The development of a full account of behavior does not require the understanding of mental events.
8. Mental events, if they do exist, are simply “way stations” between environmental stimuli and overt behaviors, and therefore can be ignored.

Not all psychologists in the early to mid-twentieth century followed the strict path of behaviorism. Wolfgang Köhler in Germany carried out studies of problem solving in chimpanzees. His findings appeared in the volume *The Mentality of Apes* in 1925. Köhler rejected the view that learning necessarily proceeded gradually or in a purely mechanical way through trial and error, as proposed by Thorndike. Instead, Köhler noted that his animals, after some unsuccessful attempts at solving a problem, might suddenly grasp the solution, displaying what he and others called *insight*. Furthermore, Köhler argued that animals can learn about the relations among stimuli, not just about the relations between stimulus and response.

N. R. F. Maier carried out studies of maze learning in rats during the late 1920s. The task required that the rats find food regardless of their starting point in the maze. The success of the rats prompted Maier to credit them with forming “spatial maps” of the maze that enabled them to find the shortest routes to food, rather than relying on previously learned associations between specific maze cues and reward. Maier described this process as “reasoning” in the rat. Köhler’s and Maier’s experiments are but two of many examples of the early resistance to the complete dismissal of the animal mind.

Along with this resistance came new research findings regarding animal behavior that could not easily be

explained within the confines of behaviorism. One of these challenges came from within the behaviorist camp itself. E. C. Tolman's studies of latent learning of mazes by rats led him to propose that animal behavior is largely purposive and goal directed, that animals form expectancies of outcomes of their behavior, and that they can learn about relations between stimuli without overt reinforcement. Tolman found that rats who had previous exposure to a maze without reinforcement performed better on that maze at a later time when reinforcement was available, than did rats who had never been exposed to the maze previously. His studies showed that rats could learn about the relations between stimuli without immediately demonstrating that knowledge through performance and without being overtly reinforced. These findings largely challenged the notion that all behavior could be explained using strict S-R associationism. They also demonstrated that overt reinforcement was not always necessary for learning to take place.

Other researchers challenged the behavioristic idea of the universality of the principles of reinforcement within S-R learning. Breland and Breland (1961) showed that the "laws of learning" were not indifferent to the particular species or task chosen for study. They found that some animals tended toward "instinctive drift" of trained behavior rather than remaining bound to either the law of effect or the principles of reinforcement. During early training of a specific behavior with food reward, the animal followed the principles of reinforcement quite readily, and the trained response remained stable and predictable. With more extended training, which should have stabilized behavior even further, however, the animals in fact exhibited less stable behavior. Their responses shifted from those that resulted in obtaining food to those that resembled the manipulation of food. For example, pigs trained to accept a coin and deposit it in a "piggy bank," began to toss the coins about and "root" them in the ground. These behaviors resembled their natural tendencies when searching for or obtaining food. Such changes in behavior had no easy explanation by the laws of reinforcement, which suggest that behavior should become more efficient with repeated reinforced practice.

The growing dissatisfaction with behaviorism as a basis for explaining complex animal behavior in the laboratory, coupled with observations from the wild of

animal behaviors that had previously been attributed only to humans (e.g., tool use by chimpanzees as described by Goodall, 1986), led to a revitalized interest among many psychologists in the mental activities of animals, including animal intelligence and the evolution of intelligence.

Several psychologists sought to develop tests that ranked the intelligence of different animal species. Harry Harlow was noted for his studies of "learning sets," or "learning how to learn." In studies published in 1949 and later, Harlow compared the ability of several species of monkey to increase their efficiency in learning to solve a class of problem. Harlow showed that as the number of problems of a particular type solved by the monkey increased, there was an improvement in the efficiency with which each new instance of the problem was solved. In some cases, new instances were solved after a single learning trial. He termed such improvement the development of a *learning set*. Other workers, following Harlow's lead, compared learning-set performance of primate and nonprimate species. The results were generally in keeping with expectations based on brain complexity or size: Monkeys showed more rapid attainment of learning efficiency than did cats, which in turn were superior to rats or to squirrels. Although the learning-set approach was seized upon by some as a way to measure comparative intelligence, later investigators showed that many extraneous variables, other than learning ability, could influence performance in such tasks, including the visual abilities of the species under study and the type of response used. Also, there was great variation of individuals within and between species. Some individual cats performed as well as or better than some individual monkeys.

A different approach to the study of the evolution of intelligence searched for discrete discontinuities in the learning abilities of selected species. In the 1960s M. E. Bitterman and colleagues began a series of studies comparing the performances of one representative species from each of four different classes of vertebrates: fish (goldfish), reptiles (turtle), birds (pigeon), and mammals (rat). The initial results showed discontinuities across these species in such tasks as "serial reversal learning," "probability learning," and "responses to shifts in reward values." For example, rats could learn to switch from one spatial stimulus to an-

other after a single nonreinforced trial, but the goldfish could not. Later work showed, however, that not all species within a given class performed the same way. Observed differences often resulted from differences in contextual variables. When rats that had been reinforced for running through a maze for a large number of pellets were suddenly switched to a much smaller reward, their running speeds decreased dramatically, even below that of rats that were originally trained on the small amount (Crespi, 1942). This phenomenon did not appear at all when different concentrations of a sucrose solution were used instead of food pellets (Flaherty, 1982). Furthermore, postulated discontinuities in learning mechanisms between different species (e.g., that rats can form S-S associations, but fish can only form S-R associations; Lowes & Bitterman, 1967; Bitterman, 1975), were often shown to be untenable in later studies using different methods (Bitterman, 1984b). The fact that observed differences between species performing similar tasks could often be explained largely in terms of differences in motivation, context, and perceptual mechanisms, led a number of researchers to suggest abandoning comparisons between diverse species in terms of their intelligence. In a seminal article, W. Hodos and C. Campbell (1969) argued that no theory of mental evolution could develop from studies of the species selected by Bitterman, as none of these species was ever an ancestor to another of the species. These authors described Bitterman's approach as reflecting the outdated *Scala naturae* or hierarchical view of Aristotle. The proper study of the evolution of intelligence, these authors stressed, requires the examination of species falling along the same branches of the evolutionary tree.

An even more extreme view suggests that, in general, comparisons of the intelligence of different species are really meaningless because all animals are the most intelligent for their ecological niche (e.g., Bailey & Bailey, 1986). Here intelligence is viewed as an adaptation carefully geared to the pressures of the niche occupied by the species. As a result, animals may manifest a composite of specialized intellectual abilities (Lockard, 1971). This "special abilities" view stands in contrast to the "general processes" view that all species share certain basic learning mechanisms, even from diverse phyla (Bitterman, 1984a; see also Mac-

phail, 1982, 1987, for the extreme view that no differences exist in intellect or learning mechanisms between different animal species). Both of these views may have merit. The universality of classical conditioning, for example, suggests that animal behavior may have roots in similar basic learning mechanisms and may then diverge into capabilities geared toward survival and proliferation in specific environmental and social niches. Additionally, those species that do not share a recent common ancestor but nevertheless share similar social systems or have experienced similar environmental pressures over the course of evolution may exhibit similar cognitive capabilities.

A further approach to the study of the evolution of intelligence sought to discover a neurological correlate of intelligence by examining the relative size (Jerison, 1973), structure (e.g., Morgane & Jacobs, 1972), and development of the brains of various species from different taxonomic groups. Traditionally, large brains have been associated with greater intelligence. However, a more accurate neurological estimate of behavioral capacity or intelligence relies on the ratio of brain size to body size (usually, volume is the measure of size used). H. Jerison (1973) uses the term *encephalization quotient* (EQ) to refer to this ratio. In general, within taxonomic groups, larger bodies require larger brains to control the increased number of somatic cells. "Excess" brain tissue beyond that required for basic somatic functions is presumably available for increased and advanced information processing (Jerison, 1985). Therefore, for a given species in a taxonomic group, the higher the ratio of brain to body size, the more excess cortical brain tissue and, presumably, the more information processing power available for intellectual tasks. Interestingly, when the EQ ratios are calculated for various species, the results seem to correlate roughly with human intuitions regarding the intelligence of different animal species. However, body volume estimates can vary widely for very large animals, especially for animals such as whales, which may feast for part of a year and fast during the remainder. Furthermore, as emphasized by L. M. Herman (1980), although EQs may accurately describe the evolution of encephalization and suggest cognitive potential, it is behavior, not structure, that must ultimately be the measure of the intellectual characteristics of a species.

CONTEMPORARY VIEWS OF ANIMAL INTELLIGENCE

Following the near demise of radical behaviorism for explaining complex animal behavior (or human behavior) and the unsuccessful attempts to scale the intelligence of various nonhuman species, the study of animal intelligence shifted largely to the study of animal "cognition." This shift followed on the heels of the general cognitive revolution occurring in many branches of the information sciences in the mid-1970s and the resultant great growth in the number and types of studies of human cognition. Some examples of studies of animal cognition are:

1. **Memory:** Pigeons (Santiago & Wright, 1984) and monkeys (Wright, Santiago, & Sands, 1984) may under certain conditions display both primacy and recency phenomena when remembering serial lists of objects. This situation is similar to results found for humans in which items early in the list (primacy effect) and late in the list (recency effect) are remembered better than items in the middle of the list. In other studies using serial lists of sounds, bottlenosed dolphins display strong recency effects (Thompson & Herman, 1977).

2. **Categorization:** Pigeons can reliably place classes of objects in different perceptual categories. For example, they can be taught to peck only at photographs containing persons or portions of a person (Herrnstein & Loveland, 1964). Also, chimpanzees have demonstrated that they can acquire conceptual categories. Savage-Rumbaugh and associates (1980) showed that common chimpanzees can sort items into "food" and "nonfood" categories where the categories are represented by abstract symbols.

3. **Reference:** In the laboratory setting chimpanzees can learn to use arbitrary symbols as references to real-world objects (Savage-Rumbaugh, 1986). Dolphins can understand a symbolic reference to absent objects (Herman & Forestell, 1985). In the wild, vervet monkeys use different vocalizations to refer to four different types of predators (Seyfarth & Cheney, 1993).

4. **Language:** Dolphins (Herman, Richards, & Wolz, 1984) can learn to understand commands given through either sequences of gestures or arbitrary computer-generated sounds, where understanding depends on

the meaning of words and word order. Similar findings have been noted for pygmy chimpanzees (bonobos) where the medium of communication is human speech (Savage-Rumbaugh et al., 1993).

5. **Counting:** A variety of different species including rats (Capaldi & Miller, 1988), pigeons (Honig & Stewart, 1989), parrots (Pepperberg, 1987); macaque monkeys (Washburn & Rumbaugh, 1991); and chimpanzees, (Boysen & Berntson, 1989; Rumbaugh, Savage-Rumbaugh, & Hegel, 1987) can learn various numerical concepts, which range from simple numerosity judgments to counting and addition.

6. **Pedagogy:** Some wild chimpanzees have shown a natural ability to "tutor" their young in the art of nutcracking (Boesch, 1991), including the skills necessary for selecting the proper tools and the method for successfully splitting the shells.

7. **Deceit:** Chimpanzees can learn to lie about the location of food when the individual human given the information will not share the food. If the human does share, then the correct location will be indicated by the chimpanzee (Woodruff & Premack, 1979). Furthermore, in the wild certain species of birds will feign injury (e.g., a broken wing) in order to lure predators away from the bird's eggs or young (Ristau, 1991).

In contrast to the way in which animals were viewed in the 1950s and 1960s by radical behaviorists, later cognitive psychologists, who primarily study animal behavior in the laboratory, and cognitive ethologists, who primarily study animal behavior in the wild, have viewed animals as active processors of information. This information comes both from the environment in the form of sensory stimulation and from the animal in the form of mental representations (Roitblat, 1982) stored in short- and long-term memory. Mental activity in animals is viewed as it is with humans, as a combination of bottom-up and top-down processing in which an animal's expectancies and previous knowledge are taken into account when new information is acquired through the senses (i.e., mental events are *not* simply way stations between stimuli and responses). Furthermore, mental events that intervene between stimulus and response can be studied through careful experimentation. Thus, memory, representation, categorization, language, reference, intentionality, strategies, planning, and deceit in animals, which were

once viewed as “taboo” among the “serious” animal behaviorists, have proliferated as topics in more recent scientific literature (e.g., Griffin, 1992; Roitblat, Herman, & Nachtigall, 1993).

How these various cognitive processes in animals relate to the concept of intelligence is still problematical. In the volume *Animal Intelligence* edited by Weiskrantz (1985), the one point of apparent agreement among the authors was that there are multiple definitions for this elusive concept. Similar controversy surrounding the definition of animal intelligence as well as its measurement appears in many of the commentaries (e.g., Hodos, 1987; Fischler, 1987) on a recent article by Macphail (1987). A closer examination of the various attempts to define animal intelligence, however, reveals several points of agreement between authors. First, flexibility or versatility of behavior is regarded as a mark of intelligence (Barlow, 1987; Griffin, 1987, 1992; Macphail, 1982). Examinations of animal intelligence should, therefore, include an exploration of the variability of behavioral strategies and performances that different species employ when faced with the diverse challenges of their world (whether that world is in nature or in the laboratory). From these data researchers can draw inferences as to the nature and characteristics of underlying cognitive processes. A useful behavioral index of flexibility is the degree to which an animal can move beyond the boundaries of what it has specifically acquired through learning or beyond the boundaries of the typical behavior for its species, perhaps even into arbitrary or symbolic realms of response.

A second point of agreement is that intelligence cannot be reduced to a single general factor but is best defined as or at least revealed by the collection of cognitive skills a species possesses. A logical extension of this view is that animal intelligence is best examined by a matrix of behavioral observations and studies to describe the “cognitive characteristics” of the species, that is, its cognitive capabilities, specializations, and limitations (Herman, 1980). This species-oriented approach has yielded new insights into the flexibility and versatility of the behaviors of several diverse animal groups: common chimpanzees, old world monkeys, particularly macaque and cercopithecus monkeys, and bottlenosed dolphins.

Taking both points of agreement together provides a method for examining intelligence within a species. This method examines behavior and cognition in both the natural and laboratory settings to provide a more complete mapping of the cognitive characteristics of the species under consideration. In the natural world, this process means documenting the types and complexities of environmental and social challenges faced by the animal, and the diversity and flexibility of responses exhibited to meet these challenges. In the laboratory setting, it often means providing the animal with novel situations involving arbitrary stimuli that test its ability to extend the boundaries of what it has been explicitly taught or that involve stimuli and behaviors that go beyond those typically found in the animal's natural world. In the latter case, the goal is to examine for cognitive traits not easily uncovered through observations in the natural world. These traits may reveal the degree of inherent flexibility of mental processes of the species. Together, the naturalistic observations and field experiments and the diversity of behavioral experiments available in the laboratory provide a richer view of mental capabilities and processes than is available through either venue alone. We can briefly illustrate this synergistic approach between field observations and laboratory experimentation through some examples of studies of the flexibility in perceiving relationships and in forming abstract concepts of a particular species, the bottlenosed dolphin (*Tursiops truncatus*).

Investigations of wild dolphin populations reveal a complex society. Individual animals form a network of social affiliations ranging from the close, long-term bonding of a mother and her calf, to the relatively favored associations between peers of similar gender and age, to the more casual short-term interactions with a variety of other herd members (Leatherwood & Reeves, 1990; Wells, 1991). The dynamic, changing pattern of associations has led some researchers to describe these populations as fission–fusion societies, in which subgroups form temporarily and then break apart, only to form again at a later time. In such societies individual animals must learn to recognize the different herd members and the social roles they play. Individuals must also learn about the social conventions of the herd, and the types of social information

exchanged among herd members. In western Australia, for example, pairs or triplets of male bottlenosed dolphins have been observed to develop close, long-term associations called coalitions (Conner, Smolker, & Richards, 1992). First-order coalitions may “capture” and herd a female for breeding purposes. However, male coalitions will sometimes ally with other such male coalitions (who are presumably at other times normally rivals) to overpower a third coalition and capture its female. These “second-order” coalitions are transient, but their formation reflects consideration of who has helped whom in the past, what other coalitions are or are not present at the time, and several other contextual variables that collectively govern the eventual decisions and strategies adopted by the first-order coalition.

Another example of the versatility of behavior is feeding. Bottlenosed dolphins have an exceptionally varied repertoire of feeding strategies, many of them opportunistic in nature: for example, following shrimp boats trawling in the Gulf of Mexico, stealing fish from baited hooks in Hawaii, and herding and driving fish onto mud banks in South Carolina. Wild bottlenosed dolphins have also cooperated in fishing activities with humans, helping to herd and concentrate schools of fish, resulting in an increased catch for both dolphin and human (Busnel, 1973). Pryor and colleagues (1990), who studied a cooperative fishery in Brazil, report that unlike other such accounts, the dolphins, not the fishermen, initiate a bout of cooperative fishing by signaling that a school of fish is present. Also, only certain dolphins participate in the fishery and apparently convey to their offspring the cooperative techniques they have developed. Such complex social relationships, dynamic in nature and highly versatile, hint at a flexible intelligence driven at least in part by recognition of different levels of relationships.

In the laboratory, several tasks have been used to examine the ability of dolphins to recognize and classify relationships, particularly arbitrary, nonsocial relationships. Studies have examined the dolphin’s ability to recognize identity relationships—that objects similar in appearance belong together. In one such task, the dolphin sees a sample stimulus and then must find that same stimulus from among two or more alternative stimuli. Typically, a limited set of stimuli trains

the animal in the task, and then researchers test generalizations about the underlying identity rule by presenting new stimuli not used during training. If the animal is capable of matching these new stimuli immediately, or nearly so, it has acquired a general solution for the task that is independent of the particular objects used. Dolphins can develop such general solutions whether the stimuli presented are arbitrary objects presented to the visual sense (Herman et al., 1989), nonbiologic sounds presented to the passive hearing sense (Herman & Gordon, 1974) or arbitrary objects presented to the active echoic sense (Herman & Pack, 1992; also see Au, 1993 for a general review of dolphin echolocation). An important fact is that dolphins can infer the general solution after only a relatively small number of training trials using only a limited number of exemplars of the rule. In at least one case (Herman, Pack & Morrel-Samuels, 1993), a dolphin developed the general solution after the very first exemplar. This stands in contrast to some other nonprimates, for example, pigeons, which also can acquire a general solution but only after thousands of trials with hundreds of exemplars (Wright et al., 1988).

More abstract tasks can test for the understanding of semantic relationships among symbols. For this type of study, a dolphin received instructions through an artificial language in which unique gestures were symbols standing for particular objects in the dolphin’s laboratory world, for specific actions that might be taken to those objects, or for constructions that could be created between two objects, such as taking one object to another (Herman, Richards, & Wolz, 1984). Which object is to be transported was governed by a syntactic rule in which the first-stated object was the destination and the second-stated object was the one to be transported. That the dolphin clearly understood the semantic relationship governed by gesture order (“To B, A take” or “To A, B take”) was shown through the appropriate reversal of its response when the sequence of object gestures was reversed, no matter what objects were referred to symbolically. Moreover, a deeper understanding of the semantic properties of objects became clear by the dolphin’s refusal to initiate a response when the object specified to be transported was in fact not transportable (for

example, it might have been an immovable object, such as a window in the dolphin's tank).

These situations are but a few examples of observations and experimentation in the field or in the laboratory. The resulting findings, together with many other findings not discussed here, extend our knowledge of the cognitive characteristics of bottlenosed dolphins and reveal the considerable degree of flexibility and versatility of behavioral responses attainable by this species. In terms of intelligence, it appears that when faced with a problem in which there is a choice between a response based on stimulus-specific learned associations and one based on a more general abstraction or global concept, the dolphin in many cases chooses the latter, more versatile strategy.

CONCLUSION

Historically, there has been a long-standing interest in animal intelligence. Intelligence in animals, like intelligence in humans, is neither easily measurable nor uncontroversial in its definition. With careful study, the intellectual traits of a species may be revealed as a complex of abilities. An apparent key trait for intelligent behavior is flexibility or versatility. In general, intelligent behavior occurs when an animal reaches beyond the boundaries of what it has explicitly learned, or beyond the typical behavior of its species into varied, arbitrary, or even symbolic realms of response.

(Preparation of this chapter was supported by Grant IBN-9121331 from the National Science Foundation and by a grant from Earthwatch.)

BIBLIOGRAPHY

- AU, W. W. L. (1993). *The sonar of dolphins*. New York: Springer-Verlag.
- BAILEY, M. B., & BAILEY, R. E. (1986). Every animal is the smartest: Intelligence and the ecological niche. In R. J. Hoage & Larry Goldman (Eds.), *Animal intelligence. Proceedings of the National Zoological Association Symposia for the Public* (pp. 105-113). Washington, DC: Smithsonian Institution Press.
- BARLOW, H. B. (1987). Efficiency, versatility, cognitive maps, and language [Commentary on *The Comparative Psychology of Intelligence*]. *Behavioral and Brain Sciences*, *10*, 657-658.
- BITTERMAN, M. E. (1975). The comparative analysis of learning. *Science*, *188*, 699-709.
- BITTERMAN, M. E. (1984a). Learning in man and other animals. In V. Sarris & A. Parducci (Eds.), *Perspectives in psychological experimentation: Towards the year 2000* (pp. 59-70), Hillsdale, NJ: Erlbaum.
- BITTERMAN, M. E. (1984b). Migration and learning in fishes. In J. D. McLeave, G. P. Arnold, J. J. Dodson, & W. H. Neill (Eds.), *Mechanisms of migration in fishes* (pp. 397-420). New York: Plenum.
- BOESCH, C. (1991). Teaching among wild chimpanzees. *Animal Behavior*, *41*, 530-532.
- BOYSEN, S. T., & BERNTSON, G. G. (1989). Numerical competence in a chimpanzee (*Pan troglodytes*). *Journal of Comparative Psychology*, *103*, 23-31.
- BRELAND, K., & BRELAND, M. (1961). The misbehavior of organisms. *American Psychologist*, *16*, 681-684.
- BUSNEL, R. G. (1973). Symbiotic relationship between man and dolphins. *Transactions of the New York Academy of Sciences*, *35*, 112-131.
- CAPALDI, E. J., & MILLER, D. J. (1988). Counting in rats: Its functional significance and the independent cognitive processes that constitute it. *Journal of Experimental Psychology: Animal Behavior Processes*, *14*, 3-17.
- CONNER, R. C., SMOLKER, R. A., & RICHARDS, A. F. (1992). Two levels of alliance formation among male bottlenose dolphins (*Tursiops sp.*). *Proceedings of the National Academy of Sciences*, *89*, 987-990.
- CRESPI, L. P. (1942). Quantitative variation of incentive and performance in the white rat. *American Journal of Psychology*, *55*, 467-517.
- DEWSBURY, D. A. (1984). *Comparative psychology in the twentieth century*. Stroudsburg, PA: Hutchinson Ross.
- FISCHLER, I. (1987). Cognitive science and comparative intelligence [Commentary on *The Comparative Psychology of Intelligence*]. *Behavioral and Brain Sciences*, *10*, 665-666.
- FLAHERTY, C. F. (1982). Incentive contrast: A review of behavioral changes following shifts in reward. *Animal Learning and Behavior*, *10*, 409-440.
- GOODALL, J. (1986). *The chimpanzees of Gombe: Patterns of behavior*. Cambridge, MA: Belknap Press.
- GRIFFIN, D. R. (1987). Phylogenetically widespread "facts-of-life" [Commentary on *The Comparative Psychology of Intelligence*]. *Behavioral and Brain Sciences*, *10*, 667-668.
- GRIFFIN, D. R. (1992). *Animal minds*. Chicago: University of Chicago Press.

- HARLOW, H. F. (1949). The formation of learning sets. *Psychological Review*, 56, 51–65.
- HERMAN, L. M. (1980). Cognitive characteristics of dolphins. In L. M. Herman (Ed.), *Cetacean behavior: Mechanisms and functions* (pp. 363–429). New York: Krieger.
- HERMAN, L. M., & FORESTELL, P. H. (1985). Reporting presence or absence of named objects by a language-trained dolphin. *Neuroscience and Biobehavioral Reviews*, 9, 667–681.
- HERMAN, L. M., & GORDON, J. A. (1974). Auditory delayed matching in the bottlenosed dolphin. *Journal of the Experimental Analysis of Behavior*, 21, 19–26.
- HERMAN, L. M., HOVANCIK, J. R., GORY, J. D., & BRADSHAW, G. L. (1989). Generalization of visual matching by a bottlenosed dolphin (*Tursiops truncatus*): Evidence for invariance of cognitive performance with visual or auditory materials. *Journal of Experimental Psychology: Animal Behavior Processes*, 15, 124–136.
- HERMAN, L. M., & PACK, A. A. (1992). Echoic-visual cross-modal recognition by a dolphin. In J. R. Thomas, R. A. Kastelein, & A. Supin (Eds.), *Sensory processes of marine mammals* (pp. 709–726). New York: Plenum.
- HERMAN, L. M., PACK, A. A., & MORREL-SAMUELS, P. (1993). Conceptual and representational abilities in bottlenosed dolphins. In H. L. Roitblat, L. M. Herman, & P. E. Nachtigall (Eds.), *Language and communication: Comparative perspectives* (pp. 403–442). Hillsdale, NJ: Erlbaum.
- HERMAN, L. M., RICHARDS, D. G., & WOLZ, J. P. (1984). Comprehension of sentences by bottlenosed dolphins. *Cognition*, 16, 129–219.
- HERRNSTEIN, R. J., & LOVELAND, D. H. (1964). Complex visual concept in the pigeon. *Science*, 146, 549–551.
- HODOS, W. (1987). Animal general intelligence: An idea ahead of its time [Commentary on *The Comparative Psychology of Intelligence*]. *Behavioral and Brain Sciences*, 10, 668.
- HODOS, W., & CAMPBELL, C. B. G. (1969). *Scala naturae*: Why there is no theory in comparative psychology. *Psychological Review*, 76, 337–350.
- HONIG, W. K., & STEWART, K. E. (1989). Discrimination of relative numerosity by pigeons. *Animal Learning and Behavior*, 17, 134–146.
- JERISON, H. J. (1973). *Evolution of the brain and intelligence*. New York: Academic Press.
- JERISON, H. J. (1985). Animal intelligence as encephalization. In L. Weiskrantz (Ed.), *Animal intelligence* (pp. 21–35). Oxford: Clarendon Press.
- KÖHLER, W. (1925). *The mentality of apes*. London: Routledge.
- LEATHERWOOD, S., & REEVES, R. R. (EDS.). (1990). *The bottlenose dolphin*. New York: Academic Press.
- LOCKARD, R. B. (1971). Reflections on the fall of comparative psychology: Is there a message for us all? *American Psychologist*, 26, 168–179.
- LOWES, G., & BITTERMAN, M. E. (1967). Reward and learning in goldfish. *Science*, 157, 455–457.
- MACPHAIL, E. M. (1982). *Brain and intelligence in vertebrates*. Oxford: Clarendon Press.
- MACPHAIL, E. M. (1987). The comparative psychology of intelligence. *Behavioral and Brain Sciences*, 10, 645–695.
- MORGAN, C. L. (1894). *An introduction to comparative psychology*. London: Walter Scott.
- MORGANE, P. J., & JACOBS, M. S. (1972). Comparative anatomy of the cetacean nervous system. In R. J. Harrison (Ed.), *Functional anatomy of marine mammals: Vol. 1* (pp. 117–244). New York: Academic Press.
- PEPPERBERG, I. M. (1987). Evidence for conceptual quantitative abilities in the African grey parrot: Labelling of cardinal sets. *Ethology*, 75, 37–61.
- PRYOR, K., LINDBERGH, J., LINDBERGH, S., & MILANO, R. (1990). A dolphin–human fishing cooperative in Brazil. *Marine Mammal Science*, 6, 77–82.
- RISTAU, C. A. (1991). Aspects of the cognitive ethology of an injury-feigning bird, the piping plover. In C. A. Ristau (Ed.), *Cognitive ethology: The minds of other animals* (pp. 91–126). Hillsdale, NJ: Erlbaum.
- ROITBLAT, H. L. (1982). The meaning of representation in animal memory. *Behavioral and Brain Sciences*, 5, 353–406.
- ROITBLAT, H. L., HERMAN, L. M., & NACHTIGALL, P. E. (EDS.). (1993). *Language and communication: Comparative perspectives*. Hillsdale, NJ: Erlbaum.
- ROMANES, G. J. (1882). *Animal intelligence*. New York: Appleton.
- RUMBAUGH, D. M., SAVAGE-RUMBAUGH, S., & HEGEL, M. T. (1987). Summation in the chimpanzee (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes*, 13, 107–115.
- SANTIAGO, H. C., & WRIGHT, A. A. (1984). Pigeon memory: Same/Different concept learning, serial probe recognition acquisition, and probe delay effects on the serial-position function. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 498–512.
- SAVAGE-RUMBAUGH, E. S. (1986). *Ape language: From conditioned response to symbol*. New York: Columbia University Press.

- SAVAGE-RUMBAUGH, E. S., MURPHY, J., SEVCIK, R. A., BRAKKE, K. E., WILLIAMS, S. L., & RUMBAUGH, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, 58 (3-4, Serial No. 233).
- SAVAGE-RUMBAUGH, E. S., RUMBAUGH, D. M., SMITH, S. T., & LAWSON, J. (1980). Reference: The linguistic essential. *Science*, 210, 922-925.
- SEYFARTH, R. M., & CHENEY, D. L. (1993). Meaning, reference, and intentionality in the natural vocalizations of monkeys. In H. L. Roitblat, L. M. Herman, & P. E. Nachtigall (Eds.), *Language and communication: Comparative perspectives* (pp. 195-219). Hillsdale, NJ: Erlbaum.
- THOMPSON, R. K. R., & HERMAN, L. M. (1977). Memory for lists of sounds by the bottlenosed dolphin: Convergence of memory processes with humans? *Science*, 195, 501-503.
- THORNDIKE, E. L. (1898). Animal intelligence: An experimental study of the associative processes in animals. *Psychological Review Monograph Supplement*, 2, 1-109.
- WASHBURN, D. A., & RUMBAUGH, D. M. (1991). Ordinal judgments of Arabic symbols by macaques (*Macaca mulatta*). *Psychological Science*, 2, 190-193.
- WATSON, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20, 158-177.
- WEISKRANTZ, L. (ED.). (1985). *Animal intelligence*. Oxford: Clarendon Press.
- WELLS, R. S. (1991). The role of long-term study in understanding the social structure of a bottlenose dolphin community. In K. Pryor & K. S. Norris (Eds.), *Dolphin societies: Discoveries and puzzles* (pp. 199-225). Berkeley and Los Angeles: University of California Press.
- WOODRUFF, G., & PREMACK, D. (1979). Intentional communication in the chimpanzee: The development of deception. *Cognition*, 7, 333-362.
- WRIGHT, A. A., COOK, R. G., RIVERA, J. J., SANDS, S. F., & DELIUS, J. D. (1988). Concept learning by pigeons: Matching-to-sample with trial-unique video picture stimuli. *Animal Learning and Behavior*, 16, 436-444.
- WRIGHT, A. A., SANTIAGO, H. C., & SANDS, S. F. (1984). Monkey memory: Same/Different concept learning, serial probe acquisition, and probe delay effects. *Journal of Experimental Psychology: Animal Behavior Processes*, 10, 513-529.

LOUIS M. HERMAN
ADAM A. PACK

ANIMAL INTELLIGENCE: PRIMATE Although intelligence is one of the oldest issues in psychology and is among the most commonly discussed and researched notions in the field, it remains characterized by the most basic of questions: What is it and how can one best measure it? These issues remain problematic for comparison of intelligence between humans of different ages, races, genders, classes, and so forth, but they are even more problematic in discussing the intelligence of organisms from different species.

Intelligence is clearly related to, but not synonymous with, learning, memory, the speed of information processing, problem solving, and language. We will define intelligence as the potential to acquire and to use knowledge. It is marked, whether between species or between representatives of a single species, by differences that are quantitative (i.e., amount of potential) and qualitative (i.e., kind of potential). Given this definition, it is clear that all primates can be characterized by some degree of intelligence. Indeed, organisms of all species can learn and master basic challenges, and thus they can be said to be intelligent to some degree. From the simplest prosimian to the most intelligent human, however, there exist differences in the speed and ease of learning, the amount that can be remembered, and the types of problems that can be solved. Assessing these differences is the crux of addressing intelligence across species.

INTELLIGENCE AND BRAIN SIZE

Before considering the strategies for assessing intelligence across species, some groundwork must be set. One means of comparing intelligence across animals involves the comparison of brain size between different species, because intelligence certainly has much to do with the brain. Indeed, there is a vast range of brain sizes across species, even if we are considering only primate species. Brain size typically varies in accordance with body size. Large-bodied species usually have larger brains than small-bodied species, and large individuals usually have larger brains than smaller individuals even within the same species. Male humans are typically larger than female humans, and males have larger brains; however, we know that, on average,

no differences exist between males and females in performance on IQ tests (Gould, 1978).

One consideration is the size of the brain relative to the size of the body as a measure. That is, one can determine for each species whether the brain is bigger or smaller than what would be predicted on the basis of body size. The most common such measure is the *encephalization quotient* (Jerison, 1976). The encephalization quotient (named for the tendency for relatively advanced behaviors and processes to be localized in the cerebral cortex) is the average brain size for a particular species compared to the average brain size for other species of the same body size. Using this measure, one can tell, for instance, that humans have disproportionately large brains, given the average height, weight, and surface skin area of our species, and that gorillas (a larger-bodied animal) are somewhat less encephalized, thus presumably less intelligent.

However, this solution takes us only partway into the issue of intelligence across species. The presumed relation between encephalization quotient (or other measure of brain size) and intelligence still requires some independent, psychological measure of intelligence. Although undeniably related, intelligence is not just brain size any more than the quality of a book is determined by the number of pages. However, few would argue that intelligence is inversely related to brain size; consequently, encephalization provides one of several means for validating other proposed intelligence measures.

INTELLIGENCE AND ENVIRONMENT

Scholars have waged great arguments about whether human intelligence is an entity fixed by heredity or a product of learning and interaction with the environment. Psychologists today believe that intelligence is the product of interactions between heredity and experience (i.e., environment). Intelligence is neither fixed nor of a finite quantity and varies in its scope and efficiency across the lifespan of an individual.

In particular, early environment and methods of rearing also can have profound effects upon intelligence. Chimpanzees reared for the first year of life in impoverished environments remained cognitively in-

ferior as adults when compared to other chimpanzees that had been group-reared during their first year of life. Chimpanzees' global learning ability (learning on a wide variety of tasks) appears to be much more vulnerable to the effects of impoverished rearing than is the ability of rhesus monkeys. This observation suggests that the intelligence of adults whose species are noted for having very complex brains, such as our own, is particularly dependent on good rearing and ample stimulation from birth.

INTELLIGENCE AND THE ANECDOTE

Early discussions about intelligence in nonhumans (dating to Romanes in 1882) relied on the anecdotal method for evidence. Anecdotes are essentially single reports of fortuitous observations. People enjoy recounting the clever actions of their pets—little problems they solve, sounds they make, and so forth. Even though some anecdotes have significant grains of truth, they are of minimal value to us because, by their very nature, they are not subject to replication or control, hence, are not scientific. Anecdotes may be useful for identifying interesting phenomena that can be studied empirically.

INTELLIGENCE AND PROBLEM SOLVING

One early attempt to examine clever behaviors by nonhuman primates in a scientific context was the classic series of experiments by Kohler (1925). In some of his best known studies, Kohler suspended a prized incentive, such as a banana, from a string so that chimpanzees could see but not reach it. Available to the chimpanzees was a variety of items—sticks, boxes, and so forth—with which the animal was familiar but which had never been used to retrieve an incentive in such a context. Would the chimpanzees' perception of the components of the problem permit them to reorganize those components so that they could stack boxes to build a tower and reach the incentive? Would they use the stick to strike it and make it fall? What kinds of ineffective tactics would be attempted? Questions of these kinds were reflections of the Gestalt psychologists, who were interested in insightful learn-



Figure 1

*Kanzi, a juvenile male bonobo chimpanzee (*Pan paniscus*), was the first of four to manifest language development in a pattern that is very similar to that of the human child—first, to comprehend and then “talk.” Kanzi talked through use of computerized keyboards, because chimpanzees cannot speak.*

ing rather than learning based on the gradual trial-and-failure method, as in the learning of mazes. Chimpanzees were typically able to view the objects as relevant to the problem, to stack the boxes and use the stick to obtain the banana, without repeated trials and failures.

Kohler's, and later Yerkes's (1929), apes were believed to have solved problems of this kind on the basis of insight into the essence of a problem and how it might be solved. It was the perceptual organizational processes of the chimpanzees, according to the Gestalt perspective, that defined new relationships between the use of these objects and the solution of the problems. Although no specific amount of intelligence was assigned to such achievements by the apes, probably everyone is inclined to accept them as manifestations of relatively advanced intelligence—adaptations afforded not by gradual trial-and-error learning, but rather by the calculations of an active, problem-oriented chimpanzee.

Kohler's view sharply contrasted with that of ardent behaviorists, whose emphasis was only on observable

behavior and not mental operations. Behaviorists objected to references to “bright” animals and “insight,” and argued, for instance, that for every lost dog that found its way home from a distance, there were large numbers that never did. Thus, only by chance circumstances did animals ever appear smart.

The contrasting perspectives of the Gestaltists, represented by Kohler, and the behaviorists in the early part of the twentieth century can be amalgamated by more recent research. The learning of specific responses to specific stimuli, in the style of the behaviorists, here will be abbreviated as stimulus-response associative learning. The problem-solving achievements of Kohler's apes will be termed relational learning, because a subject learns about the relations between stimuli, objects, and the consequences of responding to them in various ways. For both forms of learning, experience is necessary. For stimulus-response associative learning to occur, there must be repeated opportunities (i.e., trials) to learn and, characteristically, the gradual elimination of errors. For Kohler's relational learning, the chimpanzees only had to have prior experience with boxes and sticks to learn their attributes before they might be perceived as tools with which to obtain an otherwise inaccessible incentive.

INTELLIGENCE AND LEARNING

Harlow (1949) also set out to quantify learning, with the goal of undermining the notion of insight and demonstrating the importance of associative learning in problem solving. Instead, he discovered one of the important principles of relational learning and pioneered the most popular method to date for assessing intelligence in nonhuman animals. The important principle he discovered was termed learning set or learning-to-learn, and it was based on the observation that monkeys, given repeated opportunities to learn, not only learn what they are being taught but also learn to learn more efficiently. The pioneering method used to study learning set was the discrimination learning paradigm, in which subjects had to learn which of two novel stimuli was the correct one (the one associated with reward) given only six trials of experience with the material to be discriminated. Harlow observed

that, with experience, monkeys could identify the correct stimulus in fewer and fewer trials, and he reasoned that performance on the second trial of each problem (i.e., Trial 2 with each novel pair of stimuli) would constitute a good measure of intelligence.

Trial 2 learning set measures have been obtained for a variety of species, and generally correlate with measures of brain size. Unfortunately, learning set performance has since been found to be confounded with species-specific sensory-motor variables. That is, how different species perform on tests of learning set will differ as a function of visual ability, motor coordination, and a variety of other variables, as well as intelligence. Thus, discrimination learning set has generally been invalidated as a general animal intelligence test.

A variety of other learning measures have also been proposed (see Fobes & King, 1984). Most successful among these efforts has been the use of the transfer index (TI) and its derivatives (Rumbaugh & Pate, 1984). Transfer index is an extension of the two-choice discrimination learning model already discussed. In this procedure, a discrimination problem is presented to a subject until performance reaches some criterion, at which time the cue values of the materials to be discriminated are reversed for eleven trials. Rather than examining initial learning of the discrimination (as in learning set), which is biased by numerous stimulus, response, experiential, and species characteristics, the TI focuses on the transfer of training demanded by these reversal trials (i.e., how does the subject apply what it has previously learned to its performance on reversal trials?). The measure itself is a within-subject ratio of performance on reversal trials to performance on initial acquisition; by examining a subject's reversal performance relative to prereversal accuracy, the measure is relatively immune to sources of bias.

Average TI values have been reported for a variety of species. Across the primate order, TI values correlate significantly with brain development. Further, the direction (positive versus negative) and degree of transfer when acquisition criterion is changed has also been shown to vary with taxonomic status. The Spearman rank-order correlation between brain complexity and direction and degree of transfer for seven primate categories ranging from prosimian to human has been

reported to be as high as 0.93 ($p < 0.01$; Rumbaugh & Pate, 1984, p. 580). Thus species with relatively large brain-to-body size such as the great apes and humans are more likely to show generalized benefits from learning (i.e., the rule-like relational learning already discussed). In contrast, small monkeys and prosimians, whose brains tend to be small and with relatively uncomplicated cortices, are noted for a lack of transfer of learning and for stimulus-response associative, rather than relational, learning styles. Additionally, because no other learning measurement correlates more highly with achievement levels on a battery of eighteen psychological tasks than does rhesus monkeys' transfer-of-learning skills, it is suggested that there is, indeed, a global or general factor in nonhuman primate intelligence, as has been argued for human intelligence. This does not deny the additional existence of specific areas or types of intelligence (see Gardner, 1983; Sternberg & Wagner, 1986).

INTELLIGENCE AND MEMORY

Memory is implicit in all investigations of intelligence. Additionally, the duration, capacity, modality, maintenance, and other characteristics of working memory have been examined across primate species. Although individual and species differences in working memory do not alone account for variations in intelligence, it is clear that memory influences both its manifestation and its assessment across species.

INTELLIGENCE AND PROCESSING SPEED

In an attempt to measure this general intelligence factor in humans in such a way as to avoid sociocultural or other environmental sources of bias, Jensen (1982), Hunt (1983), and others have turned to the measurement of mental-processing speed as a measure of intelligence. They have reported that scoring highly on standardized assessments of intelligence is associated with relatively fast mental processing of an event (i.e., discriminating a stimulus, making a decision, and initiating a response). Thus, there is substantial support for the validity and utility of speed-of-processing assessments of general intelligence, although the matter is not without contention.

Similar correlations have been found for rhesus monkeys between intelligence, as determined in the transfer index model and speed-of-processing, measured with a choice reaction-time task. In fact, significant negative correlations (i.e., higher intelligence scores were associated with faster processing times) were most pronounced between speed-of-processing and degree of relational (versus associative) learning. Again, these data support a continuum of general intelligence across species, as well as an important qualitative difference between associative and relational learning capacity.

INTELLIGENCE AND LANGUAGE

In the past it was widely held that the key dimensions of intelligence—the ability to think and reason, the ability to use symbols to assist problem solving, and the capacity for a species to enculturate—are contingent upon language. Language was thought to be dependent upon speech, which was enabled by a mutation unique to humans. Thus, given this perspective, it was concluded that nonhuman animals had no intelligence because they had no language.

In the 1950s, there was a renewed interest in the question, Do apes have any capacity for language? Language, as used here, excludes a variety of communicative behaviors observed in nature, but rather refers to the ability to become competent in the use of even an arbitrary set of symbols (i.e., words, signs, geometric patterns, and so forth) that serve to represent things not necessarily present in time and space that could be used for the purpose of social discourse and commerce. Savage-Rumbaugh and her colleagues observed that Kanzi, then a 2½-year-old bonobo (*Pan paniscus*), learned a keyboard-based language by observation, by being reared in a language-enriched environment where he saw people and other chimpanzees use language and its effects. Several years later, and employing rigorously controlled tests, Savage-Rumbaugh and her colleagues (1993) reported that Kanzi could comprehend the syntax of newly *spoken* requests commensurate with the ability of a 2½-year-old child, and he could structure grammatical communications through use of keyboard symbols and gestures at the level of a 1½-year-old child. Their

report, based on comprehension, serves effectively to counter an earlier conclusion drawn by Terrace (1979), which stated that apes' skills of language production (i.e., manual signs) were due primarily to imitation and not competence.

To determine whether these revolutionary language skills were the result of a species-specific intelligence or a general primate intelligence coupled with a rich, language-steeped environment, a bonobo and a common chimpanzee were coreared from six weeks of age. By the age of 3 years, both evidenced comprehension of single spoken words and competent use of their word-lexigrams (e.g., geometric symbols on computer-controlled keyboards) in controlled tests. The bonobo was about three times more competent than was the common chimpanzee, however, and at age 6 could comprehend new sentences of request. These and other observations suggest an interaction between species of chimpanzee and the readiness to acquire language during infancy.

The most basic question of apes' potential for language can now be answered in the affirmative. Research with apes has served to redefine language in a substantive way, with the emphasis placed on comprehension and not speech. Although the speech and hearing systems of our species uniquely afford spectacularly efficient, distal, linguistic communication, speech should not be equated with language (or vice versa). Speech mechanisms serve only to disturb the air molecules and to send vibrations (i.e., sound waves) to potential listeners. There is nothing, however, inherently intelligent in those waves of sound. The intelligence of speech is interpreted or inferred by the listener. Hence, language is first and foremost comprehension, and not speech production. Children come to comprehend speech before beginning to talk. We have learned that if a chimpanzee is reared like a human child, it, too, will come to comprehend human speech and spontaneously come to use symbols for social communication, albeit at an early developmental level.

It is clear that language affords a remarkable economy for intelligence, permitting knowledge to be represented, encoded, stored, retrieved, transformed, and transmitted efficiently. It is also clear that relational intelligence more than species membership is prereq-

uisite for language. To hold that language is due primarily to a mutation that is unique to humans is no longer tenable. Rather, language appears to be primarily dependent on the interaction between relational intelligence and upon learning and the experiences of infancy.

BIOLOGICAL SMARTNESS AND PSYCHOLOGICAL SMARTNESS

These findings suggest two major vectors of smartness—one biological and the other psychological—that afford behavioral adaptation among animals. Biology always both enables and limits behavior, and it is recognized that many behavioral adaptations are primarily determined by heredity. Fish, birds, and insects in particular are noted for unlearned, species-specific behaviors. Their reproductive, migratory, shelter-seeking, and nest-building behaviors, for example, are highly predictable if we know enough about an animal's taxonomic classification, its age, hormonal status, the environment, and time of year. These behaviors are frequently called instinctive, and although experience can be critical to their timely appearance, they are not obviously dependent upon learning. They are genetically dictated expressions of "biological smartness," and can serve the interests of adaptation and reproduction quite well. Animals whose adaptation depends heavily upon such mechanisms have quite predictable—and typically successful—styles of life.

By contrast, "psychological smartness" provides for greater opportunity for the species both to excel in its natural habitat and also to depart from its normative life-style as well. Psychological smartness is made possible by an extraordinary elaboration of the cerebral cortex and by the organizing effects that patterns of early experiences have upon the brain. Although still both enabled by and constrained by biology, intelligence enables our species to live in its natural niche, in the world of the chimpanzee, or even in space. Similarly, psychological smartness allows the great ape either to succeed in the wild or, given sufficient early experiences, to measure up even to our anthropomorphic standards for intelligence.

[Much of the research reported here was supported by a National Institutes of Health grant (NICHD-

06016) and NASA grant NAG2-438 to Georgia State University.]

(See also: EVOLUTION OF HUMAN INTELLIGENCE.)

BIBLIOGRAPHY

- FOBES, J. L., & KING, J. E. (1982). Measuring primate learning abilities. In J. L. Fobes & J. E. King (Eds.), *Primate behavior*. New York: Academic Press.
- GARDNER, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Knopf.
- GOULD, S. J. (1978). Women's brains. *Natural History*, 89, 44–50.
- HUNT, E. (1983). On the nature of intelligence. *Science*, 219, 141–146.
- JERISON, H. J. (1985). Animal intelligence as encephalization. *Philosophical Transactions of the Royal Society of London*, B308, 21–35.
- KOHLER, W. (1925). *The mentality of apes*. New York: Kegan Paul, Trench, Trubner.
- KRASNEGOR, N. A., RUMBAUGH, D. M., SCHIEFELBUSCH, R. L., & STUDDERT-KENNEDY (1991). *Biological and behavior determinants of language development*. Hillsdale, NJ: Erlbaum.
- LEGER, D. W. (1992). *Biological foundations of behavior: An integrative approach*. New York: HarperCollins.
- MYERS, D. G. (1990). *Exploring psychology*. New York: Worth.
- RUMBAUGH, D. M., & PATE, J. L. (1984). The evolution of cognition in primates: A comparative perspective. In H. L. Roitblat, T. G. Bever, & H. S. Terrace (Eds.), *Animal cognition*. Hillsdale, NJ: Erlbaum.
- SAVAGE-RUMBAUGH, E. S., MURPHY, J., SEVCIK, R. A., BRAKKE, K. E., WILLIAMS, S. L., & RUMBAUGH, D. M. (1993). Language comprehension in ape and child. *Monographs of the Society for Research in Child Development*, 58, (3 and 4).
- SEYFARTH, D. L., & SEYFARTH, R. M. (1990). *How monkeys see the work: Inside the mind of another species*. Chicago: University of Chicago Press.
- STERNBERG, R. J., & WAGNER, R. K. (EDS.). (1986). *Practical Intelligence*. New York: Cambridge University Press.
- TERRACE, H. S. (1979). *NIM*. New York: Knopf.
- WASHBURN, D. A., & RUMBAUGH, D. M. (1991). Rhesus

monkey (*Macaca mulatta*) complex learning skills reassessed. *International Journal of Primatology*, 12 (4), 377–388.

YERKES, R. M., & YERKES, A. W. (1929). *The great apes: A study of anthropoid life*. New Haven, CT: Yale University Press.

DUANE M. RUMBAUGH
DAVID A. WASHBURN

ANXIETY The relationship between anxiety and intelligence has been of major interest to psychologists for many years. This can be seen in the work of both Lewis TERMAN and David WECHSLER, pioneers in the development of the most widely used measures of intelligence for children and adults. Although Terman's (1925) studies of genius were centrally concerned with the genetic determinants of intelligence, he recognized the potential impact of "nervousness" on intellectual functioning. On the basis of his longitudinal studies, Terman concluded that intellectually gifted children were superior in emotional stability, typically experiencing fewer nervous symptoms. (See TERMAN'S GIFTEDNESS STUDY.)

While Terman's main interest was in the emotional stability of gifted children, Wechsler (1943, 1950) focused directly on how anxiety influenced different aspects of intellectual functioning. Indeed, in his early writings, Wechsler devoted an entire chapter to the clinical measurement of anxiety (Wechsler & Hartogs, 1945). According to Wechsler (1958, pp. 175–176), anxiety was "generally disruptive or disabling" in terms of its effects on attention, concentration and immediate memory under the time pressures associated with scales he developed to measure different aspects of intelligence (Wechsler, 1939).

The stress of being tested and evaluated is especially debilitating for anxiety-prone individuals. For example, Wechsler noted that "when the neurotic does poorly on the Digit Span Test, it is not because of defective memory, but generally because of a basic anxiety mobilized by the test" (1950, p. 45). Thus Wechsler recognized the importance of distinguishing between anxiety as an emotional state and neuroticism as a personality trait. Neurotic persons are more prone to respond to stressful circumstances with intense anxiety (Spielberger, 1966).

Research has demonstrated that anxiety as an emotional state as well as individual differences in anxiety

as a personality trait may influence performance on intelligence tests (e.g., Matarazzo, 1972; Spielberger, 1958; Spielberger & Katzenmeyer, 1959). We will first review the findings of investigations of the relation between anxiety and general measures of intelligence, and will then evaluate research on the effects of anxiety on different aspects of intelligence, taking into account the conditions under which the intelligence tests are given.

THE RELATION BETWEEN ANXIETY AND INTELLIGENCE

In most of the early studies of anxiety and intelligence, the relationship between general or global measures of these constructs was examined. Since many different and varied measures of both anxiety and intelligence were employed in this research, it is not surprising that the results are inconsistent and difficult to interpret. In a number of early studies (e.g., Calvin et al., 1955; Milgram & Milgram, 1977), negative relations were reported between measures of general anxiety, such as the Taylor (1953) Manifest Anxiety Scale (TMAS), and intelligence as assessed by the academic aptitude measures that are routinely administered to college freshmen. However, other researchers using similar measures failed to find any relationship between anxiety and intelligence (e.g., Klugh & Bendig, 1955; Sarason, 1956).

Schulz and Calvin (1955) found no relation between anxiety and intelligence and concluded that the inconsistent results across studies resulted from variations in selecting subjects and in the populations tested. To evaluate the possibility that the inconsistencies in these studies resulted from sampling differences, Spielberger (1958) examined the relationship between anxiety and intelligence scores for university students enrolled in introductory psychology courses during six consecutive semesters. Although the overall correlation for his total sample of more than 1,100 students was essentially zero, significant negative correlations between anxiety and intelligence were found for those semester samples in which average intelligence scores were relatively low. In other words, high anxiety adversely affected the intelligence test scores of low-ability students while actually facilitating the performance of some of the brightest students.

The effects of stressful situational factors on the relationship between anxiety and intelligence can be seen in a study by Lewis and Adank (1975). These investigators found a negative correlation between anxiety and intelligence for fourth-, fifth-, and sixth-grade elementary schoolchildren in a traditional classroom setting, but no relationship was found for children given individualized instruction. Traditional instruction was apparently more stressful than individualized instruction, and thus aroused more anxiety, which interfered with intellectual processes.

ANXIETY AND PERFORMANCE ON THE WECHSLER INTELLIGENCE TEST SUBSCALES

A number of investigators have examined the relationship between anxiety and different components of intelligence, such as those assessed by the subscales of the Wechsler intelligence tests. This research may be divided into two general types: (1) studies of the relationship between measures of general (trait) anxiety and the Wechsler subtest scores, and (2) studies designed to test the effects of experimentally or situationally induced anxiety on intelligence test scores. While the results of studies of the relation between trait anxiety and intelligence have been inconsistent and often contradictory (e.g., Dana, 1957; Kraus, 1965; Matarazzo, 1955), persons high in trait anxiety tend to do more poorly on certain Wechsler subtests when taking the tests under stressful conditions.

Extensive research has been devoted to evaluating Wechsler's (1958) original hypothesis that anxiety disrupts memory and attention, as reflected in scores on the Wechsler Digit Span subtest. Although no relationship was found between trait anxiety and digit span in a number of studies (e.g., Lewinski, 1945; Steyaert & Snyder, 1985), Moldawsky and Moldawsky (1952) observed that digit span was especially sensitive to situational stress, whereas the Wechsler Vocabulary subtest was not. Hodges and Spielberger (1969) found a significant inverse relationship between digit span and level of state anxiety experienced while taking the test, but no relation between digit span and the TMAS, which measures trait anxiety. It would seem that situational stress adversely affects the performance of students with high trait anxiety during intelligence

tests by evoking high levels of state anxiety that interfere with memory and attention, as has been noted by Matarazzo (1972).

In studies where anxiety was experimentally induced, elevations in state anxiety are consistently associated with poorer performance on those aspects of intelligence that involve attention, concentration, and memory. For example, Walker and Spence (1965) found decrements in digit span for subjects who reported feeling anxious ("distressed") when they were told they were selected for the experiment because of questionable academic performance, but no decrements for students given the same instructions who did not report feeling distressed. Additional support for the hypothesis that situationally induced anxiety disrupts performance on the Wechsler subtests has been reported in a number of studies.

Morris and Liebert (1969) administered the TMAS and the timed subtests of the Wechsler Adult Intelligence Scale (WAIS) to university students. Half of the subjects were informed that they were being timed (situational stress), while the remaining half were timed without their knowledge. Students with high scores on items judged to measure chronic worry performed more poorly under the "timed" than under the "untimed" condition. In contrast, students with low scores on worry items actually performed better when they were timed than did low-worry subjects who were not timed. Thus the worry component of trait anxiety appears to have a debilitating influence on intellectual functioning, but only for high-worry-prone students under stressful circumstances.

THE RELATION BETWEEN STATE AND TRAIT ANXIETY AND INTELLIGENCE

Inconsistent findings in studies of the relationship between anxiety and intelligence appear to be due, in part, to the failure to distinguish between anxiety as an emotional state (S-Anxiety) and individual differences in anxiety proneness as a relatively stable personality trait (T-Anxiety) or disposition (Cattell & Scheier, 1963; Spielberger, 1966, 1972). T-Anxiety scores provide an index of the potential to experience elevations in S-Anxiety under stressful circumstances. There is no assurance that a particular situation will be perceived as personally threatening, but the inten-

sity and duration of S-Anxiety can be reliably assessed with self-report measures (Spielberger, 1983).

In one of the earliest clinical studies of the effects of anxiety on different aspects of intelligence as measured by the Wechsler subtests, Schafer (1948) observed that "the most conspicuous features of an anxiety state were impaired attention (Digit Span), a less markedly but still noticeable impaired concentration (Arithmetic) . . . and the impaired ability to plan and later check for accuracy on Block Designs and Object Assemblies" (p. 43). Martinez-Urrutia and Spielberger (1973) tested the validity of these clinical observations in a study of Spanish-speaking Puerto Rican psychiatric patients who were informed they would be given an intelligence test. Spanish adaptations of the WAIS and the *State-Trait Anxiety Inventory* (STAI; Spielberger, 1983) were then individually administered.

The STAI S-Anxiety and T-Anxiety scales were given immediately prior to the WAIS. Following each WAIS subtest, a five-item S-Anxiety scale was administered, with instructions for the patients to report "how they felt during the particular subtest they had just finished" (Martinez-Urrutia & Spielberger, 1973, p. 203). After completion of the WAIS, the S-Anxiety scale was readministered.

The pre-WAIS T-Anxiety scale was a reasonably good predictor of the level of state anxiety experienced during testing, as well as performance on most of the WAIS subtests. In contrast, the pre-WAIS S-Anxiety scores did not correlate with any of the WAIS scores. Perhaps the most important finding in this study was that the S-Anxiety scales given immediately after each WAIS subtest correlated more highly with scores on that subtest than did T-Anxiety or the other S-Anxiety measures. Thus performance on a particular WAIS subtest was best predicted by the level of state anxiety experienced by the student while working on that subtest.

CONCLUSIONS

Psychologists and educators have been interested in the effect of anxiety on intelligence test performance since the beginning of the twentieth century. In Terman's pioneering studies, intellectually gifted children

were observed to be emotionally more stable and to experience fewer nervous symptoms than other children. Even before Wechsler developed his widely used intelligence measures, he called attention to the disruptive effects of anxiety on intellectual functioning.

The development of self-report measures of anxiety in the early 1950s greatly stimulated research on anxiety and intelligence. Contradictory findings in this research resulted from a failure to distinguish between the intensity of anxiety as an emotional state (S-Anxiety) and individual differences in anxiety-proneness as a personality trait (T-Anxiety). In more recent research, significant negative relations have been reported when intelligence tests were administered under stressful circumstances, and when a wide range of intellectual ability was sampled and the average level of intelligence was relatively low.

In general, performance on intelligence tests appears to be adversely affected by elevations in S-Anxiety induced by a variety of stressors such as timed tests, stressful instructions, and feedback indicating a negative evaluation of the individual. High levels of state anxiety experienced during tests have the most debilitating effects on worry-prone persons, and interfere with performance on those components of intelligence requiring attention, concentration, memory, and cognitive processing of complex information.

BIBLIOGRAPHY

- CALLENS, C. J., & MELTZER, M. L. (1969). Effect of intelligence, anxiety, and diagnosis on arithmetic and digit span performance on the WAIS. *Journal of Consulting and Clinical Psychology, 33*, 630.
- CALVIN, A. D., KOONS, P. B., BINGHAM, J. L., & FINK, H. H. (1955). A further investigation of the relationship between manifest anxiety and intelligence. *Journal of Consulting Psychology, 19*, 280-282.
- CATTELL, R. B., & SCHEIER, I. H. (1963). *Handbook for the IPAT Anxiety Scale* (2nd ed.). Champaign, IL: Institute for Personality and Ability Testing.
- DANA, R. H. (1957). Manifest anxiety, intelligence, and psychopathology. *Journal of Consulting Psychology, 21*, 38-40.
- HODGES, W. E., & SPIELBERGER, C. D. (1969). Digit span: An indicant of trait or state anxiety? *Journal of Consulting and Clinical Psychology, 33*, 430-433.

- KLUGH, H. E., & BENDIG, A. W. (1955). The Manifest Anxiety and ACE scales and college achievement. *Journal of Consulting Psychology, 19*, 487.
- KRAUS, J. (1965). Cattell Anxiety Scale scores and WAIS attainment in three groups of psychiatric patients. *Australian Journal of Psychology, 17*, 229–232.
- LEWINSKI, R. J. (1945). The psychometric pattern: Anxiety and neuroses. *Journal of Clinical Psychology, 1*, 214–221.
- LEWIS, J., & ADANK, R. (1975). Intercorrelations among measures of intelligence, achievement, self-esteem, and anxiety in two groups of elementary school pupils exposed to two different models of instruction. *Educational and Psychological Measurement, 35*, 499–501.
- MARTINEZ-URRUTIA, A., & SPIELBERGER, C. D. (1973). The relationship between state–trait anxiety and intelligence in Puerto Rican psychiatric patients. *Interamerican Journal of Psychology, 7*, 199–214.
- MATARAZZO, J. D. (1955). The relationship of manifest anxiety to Wechsler-Bellevue subtests performance. *Journal of Consulting Psychology, 19*, 318.
- MATARAZZO, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligent* (5th enlarged, ed.). Baltimore, MD: Williams & Wilkins.
- MILGRAM, R. M., & MILGRAM, N. A. (1977). The effect of test content and context on the anxiety–intelligence relationship. *Journal of Genetic Psychology, 130*, 121–127.
- MOLDAWSKY, S., & MOLDAWSKY, P. C. (1952). Digit span as an anxiety indicator. *Journal of Consulting Psychology, 16*, 115–118.
- MORRIS, L. W., & LIEBERT, R. M. (1969). Effects of anxiety on timed and untimed intelligence tests: Another look. *Journal of Consulting and Clinical Psychology, 33*, 240–244.
- PYKE, S., & AGNEW, N. MCK. (1963). Digit span performance as a function of noxious stimulation. *Journal of Consulting Psychology, 27*, 281.
- SARASON, I. G. (1956). The relationship of anxiety and “lack of defensiveness” to intellectual performance. *Journal of Consulting Psychology, 20*, 220–222.
- SARASON, I. G., & MINARD, J. (1962). Test anxiety, experimental instructions, and the Wechsler Adult Intelligence Scale. *Journal of Educational Psychology, 53*, 299–302.
- SCHAFFER, R. (1948). *Clinical application of psychological tests*. New York: International Universities Press.
- SCHULZ, R. E., & CALVIN, A. D. (1955). A failure to replicate the findings of a negative correlation between manifest anxiety and ACE scores. *Journal of Consulting Psychology, 19*, 223–224.
- SPIELBERGER, C. D. (1958). On the relationship between anxiety and intelligence. *Journal of Consulting Psychology, 22*, 220–224.
- SPIELBERGER, C. D. (1966). Theory and research on anxiety. In C. D. Spielberger (Ed.), *Anxiety and behavior*. New York: Academic Press.
- SPIELBERGER, C. D. (ED.). (1972). Anxiety as an emotional state. In C. D. Spielberger (Ed.), *Anxiety: Current trends in theory and research* (Vol. 1). New York: Academic Press.
- SPIELBERGER, C. D. (1983). *Manual for the State–Trait Anxiety Inventory* (Form Y). Palo Alto, CA: Consulting Psychologists Press.
- SPIELBERGER, C. D., & KATZENMEYER, W. G. (1959). Manifest anxiety, intelligence, and college grades. *Journal of Consulting Psychology, 23*, 278.
- STEYAERT, J. P., & SNYDER, J. F. (1985). Seating arrangement and state anxiety as related to performance on Digit Span and Digit Symbol of the Wechsler Adult Intelligence Scale. *Psychological Reports, 57*, 807–812.
- TAYLOR, J. A. (1953). A personality scale of manifest anxiety. *Journal of Abnormal and Social Psychology, 48*, 285–290.
- TERMAN, L. M. (1925). *Genetic studies of genius* (Vol. 1). Stanford, CA: Stanford University Press.
- WALKER, R. E., SANNITO, T. C., & FIRETTO, A. C. (1970). The effect of subjectively reported anxiety on intelligence test performance. *Psychology in the Schools, 7*, 241–243.
- WALKER, R. E., & SPENCE, J. T. (1965). Relationship between digit span and anxiety. *Journal of Consulting Psychology, 28*, 220–223.
- WECHSLER, D. (1939). *Measurement of Adult Intelligence* (1st ed.). Baltimore, MD: Williams & Wilkins.
- WECHSLER, D. (1943). Non-intellective factors in general intelligence. *Journal of Abnormal and Social Psychology, 38*, 101–103.
- WECHSLER, D. (1950). Cognitive, conative, and non-intellective intelligence. *American Psychologist, 5*, 78–83.
- WECHSLER, D. (1958). *The measurement and appraisal of adult intelligence*. Baltimore, MD: Williams & Wilkins.
- WECHSLER, D., & HARTOGS, R. (1945). The clinical measurement of anxiety: An experimental approach. *Psychiatric Quarterly, 19*, 618–635.

CHARLES D. SPIELBERGER
SUMNER J. SYDEMAN

APHASIA The loss of language affects all aspects of a person's life (Sarno, 1991). When one has had an intact language system, and neurological damage to the language system occurs, the resultant disorder is *aphasia* (Darley, 1982; Goodglass & Kaplan, 1983). Most often, the cause of this brain damage is one of several types of stroke (cerebrovascular accident, or CVA). Further, when aphasia occurs, it is generally assumed that the insult is to the left side of the brain, since this is where the majority of language processing and production abilities are located. Specifically, the location of the lesion (damage) within the left hemisphere will affect the characteristics of the resulting language deficits. Because of this, there is variability in the language profiles of individuals with aphasia. Based on groupings of these language behaviors, seven categories of aphasia have been identified, representing profiles of dimensions of both verbal output and of comprehension.

Our understanding of the neural bases of language disorders has a long history. Until the early 1860s, a number of researchers, such as Alexander Crichton, Jean-Baptiste Bouillaud, Franz-Joseph Gall, and Gabriel Andral, contributed to the discussion related to the nature of neural processes and possible brain-language relationships; however, it was the work of Paul Broca that gave initial specificity to the study of aphasia. Through postmortem examination of a patient who had limited ability to recall words, Broca was able to identify an underlying lesion in the left inferior frontal lobe of the brain. Broca referred to this language disorder as *aphemia*; today it is known as *Broca's aphasia*, and it continues to be characterized by a non-fluent expressive deficit. Thus, 1865 marked the beginning of the notion that language was housed within the left hemisphere. Not long after, in 1874, Carl Wernicke identified a language disorder due to a lesion in the left superior temporal lobe. This aphasia, characterized by fluency and limited comprehension, is often termed *sensory aphasia*, or *Wernicke's aphasia*.

Researchers have elaborated on earlier, more simplistic localizationist theories of aphasia to include broader network models (Mesulam, 1990) and have begun to appreciate the effects of more generalized brain damage, such as memory and attentional deficits, on the behavior of aphasics (McNeil, 1982). This arti-

cle provides a description of behaviors associated with aphasia, general categorizations of aphasia, classification of subtypes of aphasia, and possible concerns in the study of aphasia.

APHASIC BEHAVIORS

Agrammatism. A nonfluent speech pattern exhibited in certain subtypes of aphasia. The sound of agrammatic speech is much like the sound of a telegram; the small words in the sentence (or function words) are left out.

Anomia. Difficulty in finding words. This is a behavior that is often present in aphasia. People with anomia know the concept behind a word; however, they may be unable to retrieve the name. When patients have difficulty with a particular word, they may circumlocute by giving a description of the item or by describing its function. This technique may help a patient to get the message across without recalling the word, or may actually serve as a cue to retrieve the word. There are times when anomia is the primary language deficit. In such an instance, the diagnosis of anomic aphasia may be made.

Jargon. Jargon is jumbled-up speech. It is fluent and well inflected, so from a distance the person with aphasia may sound like he or she is speaking normally; however, the speech output is primarily made up of various types of paraphasia (see next paragraph) and is generally incomprehensible. There are persons with aphasia whose entire verbal output is made up of jargon.

Paraphasia. All types of paraphasia involve errors in the production of words that are not due to deficits in articulation (Goodglass & Kaplan, 1983). The incorrectly produced words may occur in relative isolation in otherwise fluent speech, or, in the case of jargon, they may dominate the verbal output of the person with aphasia. There are three types.

Verbal paraphasias (semantic paraphasias) include errors in which one word is substituted for another. These word substitutions may be either related or unrelated to the correct word. For example, the individual with aphasia may say the semantically related word "table" instead of "chair" or "fork" instead of "spoon"

in a related verbal aphasia. The substitution of “car” for “apple” would exemplify an unrelated verbal paraphasia.

A second classification of paraphasia involves errors in the sounds of a word. These *literal paraphasias* (phonemic paraphasias) may take the form of sound omissions (“kip” for “clip”), sound substitutions (“boon” for “book”), or a rearrangement of the sounds in the word (“umrebla” instead of “umbrella”).

In the most extreme form of paraphasia, over half the sounds in the intended word are incorrect, yielding a new word. These *neologistic paraphasias*, or *neologisms*, do not appear to be words in the language of the speaker. Examples of neologisms would include “clagna” for “pencil” or “bodel” for “hat.” Neologisms are pronounced and stressed just as if they were actual words and may not be perceived as errors by the aphasic speaker.

Stereotypes. In essence, stereotypes (speech automatisms) are verbal utterances used over and over again by an aphasic. In the case of a severe aphasia, one or two stereotypes may be the entire range of the person’s verbal output. Stereotypes have been described as being either nondictionary verbal forms (for example, “daloo daloo”) or dictionary forms (for example, “fine fine” or “that’s the way”) (Alajouanine, 1956). Blanken, Wallesch, and Papagno (1990) proposed that these speech automatisms relate to speech output alone and do not necessarily indicate the presence of severe comprehension deficits.

GENERAL CLASSIFICATIONS OF APHASIA

Fluent Aphasia Versus Nonfluent Aphasia.

The distinction between fluent and nonfluent aphasia is based on the verbal output of the aphasic individual. Fluent aphasics demonstrate normal prosodic patterns in their speech. In other words, the melodic line of speech remains intact. Nonfluent aphasics’ speech lacks normal stress and intonational patterns (Brookshire, 1992). All of the subcategories of aphasia can be described as either fluent or nonfluent. Thus the decision as to whether a patient exhibits a fluent aphasia or a nonfluent one is the first step in classifying an aphasia. The major fluent aphasias include Wernicke’s,

transcortical sensory, conduction, and anomic. Primary nonfluent aphasia types include Broca’s, transcortical motor, and global.

Posterior Versus Anterior Aphasia. With neurophysiology and lesion site in mind, some speak of aphasia as being of either an anterior type or a posterior type (Davis, 1993). This description reflects the fact that in most cases an anterior lesion (in the inferior posterior frontal lobe) results in a nonfluent type of aphasia and a posterior lesion (superior temporal lobe) results in a fluent aphasia. Thus, by using the terms *anterior* or *posterior*, there is a broad reference to the lesion area and the most common type of resulting aphasia.

TYPES OF APHASIA

Due to the great variability in aphasic behaviors, general terms, such as those represented by the dimensions of fluency or location, are very limited in providing a description of the aphasic’s language characteristics. However, through testing, behaviors can be isolated and a type of aphasia diagnosed that reflects some greater specificity of language behaviors. Diagnosis, however, is often approached with caution, since nosological terminology is meant only to provide a descriptor of relative areas of strength and weakness. Following initial classification, each patient should be viewed as an individual with a unique communicative deficit, then treated based on his or her own needs. The following classification system, based on the work of Goodglass and Kaplan (1983), is commonly used by those working in the field of aphasia. It is not presented as a sole system or complete description. For further investigation into issues surrounding classification of aphasia, the reader is referred to the works of Darley (1982), Kertesz (1979), LaPointe (1990), McNeil (1982), and Sarno (1991).

The first two types of aphasia to be described, Broca’s and Wernicke’s, have been referred to as *central aphasias* (Benton, 1991). These aphasias arise due to damage in the central portions of two primary language areas and directly affect the production and comprehension of language. Other types of aphasia are *paracentral* and may be caused by interference in the transmission of language impulses between the lan-

guage centers themselves or between language centers and other neural centers.

Broca's Aphasia (Motor Aphasia, Expressive Aphasia). Broca's aphasia is a primary nonfluent aphasia. Verbal output is characterized by halting, effortful, agrammatic utterances. Repetition of utterances may be in a range from poor to mildly deficient. Naming abilities also may fall into a range of ability from marked to mildly deficient. A key to diagnosis is poor verbal output despite fairly well-preserved auditory comprehension. Patients may appear quite disordered when they speak; they will understand much of what is said to them, however. They tend to be aware of their disorder and often become extremely frustrated by their communication deficit. Typically, the brain damage that results in this form of aphasia is located in the inferior posterior frontal lobe of the left hemisphere (Broca's area).

Wernicke's Aphasia (Sensory Aphasia, Receptive Aphasia). Wernicke's aphasia can be viewed as the opposite of Broca's aphasia. The Wernicke's aphasic patient speaks fluently; however, auditory comprehension is markedly impaired. This is not to suggest that verbal output is functional. Wernicke's aphasics may exhibit behaviors such as semantic and literal paraphasias, neologisms, and jargon. The resulting speech is often inadequate in content for effective communication. The patient with Wernicke's aphasia may be relatively unaware of his or her language deficit and may be quite verbose. The brain damage that produces this type of aphasia is commonly centered in the superior part of the left temporal lobe (Wernicke's area), just posterior to the primary area for hearing.

Global Aphasia. Aphasia in its most devastating form is *global aphasia*. Patients may not be able to produce meaningful speech or understand what is said to them. Stereotypes are most common in this population, with verbal utterances limited to the reiteration of a few words or sounds. Communication may be facilitated when it centers on topics that are of immediate relevance to the patient and when only yes-or-no responses are required. Brain damage in these cases may be extensive, involving both anterior and posterior language areas of the left hemisphere.

Transcortical Motor Aphasia (Dynamic Aphasia). This type is similar to Broca's aphasia. One of the major differences is that patients with transcortical motor aphasia are able to repeat phrases much better than those with Broca's aphasia. They may be able to repeat multiword utterances fluently, even though their self-initiated utterances may be halting and labored. The lesion that has been associated with transcortical motor aphasia is most likely in the area around Broca's area. Further, the lesion is typically smaller than that associated with Broca's aphasia.

Transcortical Sensory Aphasia (Amnesic Aphasia). Like the Wernicke's aphasic, those with transcortical sensory aphasia have difficulty comprehending the speech of others. They usually also have fluent spontaneous speech that contains paraphasias and neologisms. There is a general lack of information conveyed in verbal output. A key diagnostic variable, however, is the superior repetition skill of this group. A patient with transcortical sensory aphasia may be able to repeat long utterances and may correctly produce learned materials such as prayers or rhymes. Such lengthy coherent output may be deceiving, suggesting an intact verbal output mode. The brain damage that produces a transcortical sensory aphasia is thought to be one that isolates the language areas from the areas of the brain that interpret and integrate the language information. Thus, the language system is able to receive speech, process it, and produce a repetition, with no real "understanding" of the message.

Conduction Aphasia (Central Aphasia, Afferent Motor Aphasia). Conduction aphasia may be viewed as a breakdown in the transmission of information from the language comprehension and language formulation area (Wernicke's area) to the language output center (Broca's area). These patients may be fluent in their speech, often demonstrating well-formed lengthy utterances; however, there does appear to be a deficit related to the sequencing of the sounds of speech (phoneme sequencing). Some patients may demonstrate literal paraphasias as a result of this production deficit. When asked to repeat utterances, these patients may appear more like a patient with Broca's aphasia, exhibiting significant problems. Otherwise, they often have relatively preserved auditory comprehension and at least moderate verbal output abilities.

Anomic Aphasia (Nominal Aphasia). *Anomia* refers to difficulty finding words. In the case of anomic aphasia, the primary deficit is the retrieval of words. This type is considered the least severe of the fluent aphasias. Though a specific word may not be recalled, anomic aphasics may talk around the problem word (or use circumlocution) and even describe the object or event. Though these patients may have good understanding of the speech of others, they may fail to comprehend the words that are difficult for them to recall. The location of the lesion that produces an anomia is not well defined. This may be due to the occurrence of anomia in a variety of aphasia categories and in more generalized brain disorders, such as dementia.

FUTURE DIRECTIONS IN APHASIA

There is a great deal of interpersonal variability in the manifestation of the aphasic condition. This variability may be due, in part, to differences in the neural representation of language between individuals. Prediction of behaviors based on site of lesion must, therefore, be made tentatively. There is also variability within the behaviors exhibited by an individual aphasia patient (McNeil, 1982). Such intrapersonal variability makes isolation of an individual's true level of ability quite difficult. Currently, researchers are attempting to describe further the neural basis of language behaviors. Also, the roles of memory and attention in aphasia are being examined. With the knowledge gained from continued research, the variability both between and within aphasic patients may be better understood and addressed clinically.

Modern neuro-imaging techniques have been helpful tools in this investigation of language behaviors. Electrical stimulation of cortical tissue has helped to document individual variability of the neural representation of language (Ojemann, 1983). Computerized tomography (CT) and magnetic resonance imaging (MRI) have allowed direct observation of aphasia-producing lesions, thereby providing knowledge about the language role of specific neural regions (Basso et al., 1985; Lefkowitz & Netsell, 1993). Event-related potentials (ERPs) have been used to investigate the nature of language processing (Fischler, 1990; Hillyard &

Picton, 1987). Further, ERPs have been used directly to measure hemispheric activation following aphasia (Selinger & Prescott, 1993) and aphasics' attentional abilities (Peach, Newhoff, & Rubin, 1993). Positron emission tomography (PET) has also been important in providing an understanding of neural activity associated with language (Maziotta et al., 1982; Peterson et al., 1988). With the availability of these modern techniques, more answers regarding the nature of aphasia may be in our future.

BIBLIOGRAPHY

- ALAJOUANINE, M. S. (1956). Verbal realization in aphasia. *Brain*, 79, 1-28.
- BASSO, A., LECOURS, A. R., MORASCHINI, S., & VANIER, M. (1985). Anatomoclinical correlations of the aphasias as defined through computerized tomography: Exceptions. *Brain and Language*, 26, 201-229.
- BENTON, A. (1991). Aphasia: Historical perspectives. In M. T. Sarno (Ed.), *Acquired aphasia* (2nd ed., pp. 1-26). San Diego, CA: Academic Press.
- BLANKEN, G., WALLECH, C. W., & PAPAGNO, C. (1990). Dissociations of language functions in aphasics with speech automatisms (recurring utterances). *Cortex*, 26, 41-63.
- BROOKSHIRE, R. H. (1992). *An introduction to neurogenic communication disorders* (4th ed.). St. Louis, MO: Mosby-Year Book.
- DAMASIO, A. (1991). Signs of aphasia. In M. T. Sarno (Ed.), *Acquired aphasia* (2nd ed., pp. 27-43). San Diego, CA: Academic Press.
- DARLEY, F. L. (1982). *Aphasia*. Philadelphia: W. B. Saunders.
- DAVIS, G. A. (1993). *A survey of adult aphasia and related language disorders* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- FISCHLER, I. S. (1990). Comprehending language with event-related potentials. In J. W. Rohrbaugh, R. Parasuraman, & R. Johnson, Jr. (Eds.), *Event-related brain potentials: Basic issues and applications* (pp. 165-177). New York: Oxford University Press.
- GOODGLASS, H., & KAPLAN, E. (1983). *The assessment of aphasia and related disorders* (2nd ed.). Philadelphia: Lea & Febiger.
- HILLYARD, S. A., & PICTON, T. W. (1987). Electrophysiology of cognition. In S. R. Geiger (Ed.), *Handbook of physiology* (pp. 519-584). Bethesda, MD: American Physiological Society.

- KERTESZ, A. (1979). *Aphasia and associated disorders: Taxonomy, localization, and recovery*. Orlando, FL: Grune and Stratton.
- LAPOINTE, L. L. (1990). *Aphasia and related neurogenic language disorders*. New York: Thieme Medical Publishers.
- LEFKOWITZ, D., & NETSELL, R. (1993). Neuroanatomy of speech: An MRI atlas. *Journal of Medical Speech-Language Pathology, 1*, 3–26.
- MAZIOTTA, J. C., PHELPS, M. E., CARSON, R. E., & KUHL, D. E. (1982). Tomographic mapping of human cerebral metabolism: Auditory stimulation. *Neurology, 32*, 921–937.
- MCNEIL, M. R. (1982). The nature of aphasia in adults. In N. J. Lass, L. V. McReynolds, J. L. Northern, & D. E. Yoder (Eds.), *Speech, language, and hearing: Vol. II. Pathologies of speech and language* (pp. 692–740). Philadelphia: W. B. Saunders.
- MESULAM, M. M. (1990). Large-scale neurocognitive networks and distributed processing for attention, language, and memory. *Annals of Neurology, 28*, 597–613.
- OJEMANN, G. A. (1983). Brain organization for language from the perspective of electrical stimulation mapping. *The Behavioral and Brain Sciences, 2*, 189–230.
- PEACH, R. K., NEWHOFF, M., & RUBIN, S. S. (1993). Attention in aphasia as revealed by event-related potentials: A preliminary investigation. In M. L. Lemme (Ed.), *Clinical Aphasiology* (Vol. 21, pp. 311–322). Austin, TX: PRO-ED.
- PETERSON, S. E., FOX, P. T., POSNER, M. I., MINTUN, M. A., & RAICHEL, M. E. (1988). PET studies of the cortical anatomy of single-word processing. *Nature, 331*, 585–589.
- SARNO, M. T. (1991). *Acquired aphasia* (2nd ed.). San Diego, CA: Academic Press.
- SELINGER, M., & PRESCOTT, T. E. (1993). Cortical-evoked potentials and aphasia: A follow-up case study. In M. L. Lemme (Ed.), *Clinical Aphasiology* (Vol. 21, pp. 311–322). Austin, TX: PRO-ED.

SCOTT S. RUBIN
GEORGE W. HYND

APTITUDE TESTS Forecasting success in school, training, or a career has been an important application of psychological tests since their inception in the early 1900s. For this prediction task, psychologists have focused on the role of aptitude. Essentially, aptitude is potential ability or the capacity to learn.

For example, mechanical aptitude is the capacity to acquire mechanical skills, musical aptitude is the ability to learn musical skills, and scholastic aptitude is the capacity to learn what is taught in a typical school curriculum. Although psychologists acknowledge that interest, drive, and other personality traits are important, aptitude is seen as the necessary ingredient in school, training, or career success.

Aptitude tests provide quantitative estimates of a person's potential for learning the knowledge and skills needed for school, training, or career success. The purpose of this brief commentary is to survey widely used aptitude tests in relation to specific aptitudes, schooling, and employment. This article will concentrate on the multiple aptitude batteries used in school, military, and vocational settings, but also make brief reference to special aptitude tests. In addition, it will discuss some promising developments in aptitude tests and highlight certain cautions and concerns about their use.

Most aptitude tests are paper-and-pencil measures suitable for group testing. The few apparent exceptions arise from inconsistent use of terminology. For example, the Detroit Tests of Learning Aptitude–2, or DTLA–2 (Hammill, 1985), must be administered one-on-one by a trained clinician. In spite of its title, most psychologists consider the DTLA–2 to be an individual intelligence test. Another test with an arguably misleading title is the Multidimensional Aptitude Battery, or MAB (Jackson, 1984). Even though it is a paper-and-pencil measure suitable for group testing, the MAB is clearly an intelligence test. In fact, the MAB was purposely designed, subtest by subtest, to mirror the popular Wechsler Adult Intelligence Scale–Revised (Wechsler, 1981).

The distinction between aptitude, intelligence, and achievement tests is important in this context. In general, intelligence tests sample a broad assortment of cognitive skills in order to estimate current intellectual functioning. Achievement tests measure current skill attainment in relation to identified educational goals of school or training programs. In contrast, aptitude tests assess several distinctive segments of ability (in the case of multiple aptitude batteries) or a single focused ability (in the case of special aptitude tests) for the purpose of predicting future performance.

Actually, the distinction between these three types of tests is often fuzzy (Gregory, 1992). The correlations among scores on aptitude, intelligence, and achievement tests can be substantial, and items from these tests may be similar in style and content. In large measure, aptitude tests are defined by their practical applications. The most important of these applications are gatekeeping functions, such as school admission, corporate employment, and military entry, and guidance functions, such as academic counseling, career advising, and job placement. These uses of aptitude tests are discussed below.

COLLEGE ADMISSION TESTS

An important use of aptitude tests is the prediction of performance in college. Based on the knowledge that low scores on a scholastic aptitude test predict college failure, many institutions of higher education set minimum standards for admission. The two most widely used tests for this purpose are the SCHOLASTIC ASSESSMENT TESTS (SAT) and the AMERICAN COLLEGE TEST (ACT) Assessment Program.

The oldest of the college admissions tests is the SAT, which dates back to 1926. The SAT is updated on a yearly basis. The test yields separate verbal and mathematics scores, reported on a scale that ranges from 200 to 800. Characteristic item types for the verbal portion include the following (Gregory, 1992):

- Antonyms: Choose the word that is most nearly opposite in meaning to a given word.
- Analogies: Select a pair of words that best expresses a relationship similar to that expressed in a stimulus pair.
- Sentence completions: For a sentence with one or two blanks, choose a word or pair of words that best fits the meaning of the sentence as a whole.
- Reading comprehension: Read a passage and answer multiple-choice questions based on what is stated or implied in that passage.

Characteristic item types for the mathematics portion include the following (Gregory, 1992):

- Regular mathematics: Solve basic problems in geometry and algebra.
- Quantitative comparisons: Choose from two quantities which is greater, or denote that they are equal, or denote that the problem is unsolvable from the information given.

A recurring misconception about the SAT is that the mean score each year for the verbal and the mathematics subtests is 500 and the standard deviation is 100. In fact, the mean of 500 and the standard deviation of 100 refer to the mean and standard deviation of the anchor group of 10,654 students who took the verbal portion of the SAT in 1941 (Donlon, 1984). Each year, current scores are equated to the anchor scores by linking each new form of the test to a previous form. Since some test items are always carried forward to the next year, it is possible to compare the relative difficulty of new forms with old ones. This procedure guarantees that yearly SAT scores are based on the same measurement scale used in 1941. Thus, it is possible to chart long-term national trends in SAT scores, even though the specific items on the test are revised and updated on a regular basis.

For reasons that are not fully understood, mean SAT scores declined significantly from 1962 to the mid-1980s and then began an upward turn. This was especially true for the verbal portion of the test. In 1983 the mean SAT verbal score reached a low of 420, almost a standard deviation below the results for the 1941 normative sample. Some persons have speculated that social changes such as the expansion of television may have contributed to the decline. Another hypothesis is that the scholastic aptitude of America's youth has changed little over the decades, but that the sample of students enticed to take the SAT in the 1970s and 1980s included a larger proportion of academically less capable college aspirants, resulting in an artificial decline of average SAT scores. As reviewed by T. F. Donlon (1984), these hypotheses are difficult to verify.

The SAT meets high standards of technical excellence. The verbal and mathematics subsections each

TABLE 1
Subtests and item types for the ACT

<i>Subtest</i>	<i>Item Types</i>
English usage (75 items, 40 minutes)	Choose revisions for short prose passages.
Mathematics usage (40 items, 50 minutes)	Solve basic college mathematics problems.
Social studies reading (52 items, 35 minutes)	Comprehend social science writings.
Natural sciences reading (52 items, 35 minutes)	Comprehend science writings.

show test–retest reliability coefficients in the high .80s, which means that the standard error of measurement is a tolerable 30–35 points. In other words, the chances are roughly two out of three that a person’s “true score”—the score the person would get if he or she took the test over and over again—is within 30–35 points of the observed score. The ability of the test to predict college performance has been examined in more than 600 studies (Donlon, 1984). The combined SAT scores (verbal plus mathematics) correlated .42, on average, with college freshmen grade point average, but this correlation is artificially low because of the restricted range of measurement. The real correlation is no doubt substantially higher, since many students who obtain low SAT scores—and who would have received correspondingly low college grades—never attend college.

Another widely used college admission test is the ACT, initiated in 1959. The ACT consists of separate tests in English, mathematics, social studies, and natural sciences that emphasize the application of knowledge (American College Testing Program, 1988). Subtests and item types for the ACT are described in Table 1.

ACT scores are reported on a 36-point scale, with a mean of approximately 16 and a standard deviation of about 5. A composite score, which is the average of the four tests, is also provided. The ACT rivals the SAT for technical adequacy and predictive validity.

ADMISSION TO POSTGRADUATE PROGRAMS

Aptitude tests play a major role in admission to postgraduate programs such as graduate school, law school, and medical school. Indeed, it is not unusual

for graduate school admission committees to place a greater emphasis on Graduate Record Exam (GRE) scores than on any other single factor. In addition to the GRE, tests commonly used for admission to postgraduate programs are the Law School Admission Test (LSAT), required of applicants to law school, and the Medical College Admissions Test (MCAT), required of applicants to medical school. Each of these tests is discussed briefly below.

The GRE is an objectively scored group test widely used in the selection of candidates for graduate programs. In addition to advanced achievement tests in many academic fields, the GRE consists of three general sections: verbal, quantitative, and analytical aptitude. The verbal section (GRE–V) consists of verbal analogies, antonyms, sentence completion, and reading comprehension. The quantitative section (GRE–Q) contains problems in algebra, geometry, quantitative comparisons, reasoning, and the interpretation of data, graphs, and diagrams. The analytical section (GRE–A), added in 1977, consists of analytical and logical reasoning problems. All GRE scores are reported as standard scores with a mean of 500 and a standard deviation of 100. Test results are anchored to a fixed reference group of 2,095 college seniors examined in 1952 on the verbal and quantitative portions of the test.

The general GRE tests possess respectable RELIABILITY, with internal consistency reliability coefficients typically around .90 for the three components. VALIDITY is more difficult to evaluate because of the restriction of range problem. Specifically, persons with low GRE scores are rarely admitted for graduate study, so the full range of outcomes on the predictor (GRE scores) and the criterion (performance in graduate school) is unavailable for statistical analysis. Nonetheless, the combination of GRE verbal and quantitative

scores correlates around .3 with graduate school grades; the correlation is even higher, around .4, when undergraduate grades are included (Cohn, 1985; Jaeger, 1985).

The LSAT is required of applicants to virtually every law school in the United States. The LSAT consists of objective items that tap reading comprehension, logical reasoning, and analytical reasoning. These results are summarized as a single score that can range from 10 to 48 (mean of 30.5 and standard deviation of 8). An unscored writing sample is also obtained. Both the total score and the writing sample are forwarded to law schools specified by the examinee. A unique feature of the LSAT is that the items do not directly test prior knowledge of data or facts in any field. Instead, the test items are designed to tap the thinking skills needed in law school. The test has acceptable reliability (internal consistency coefficients in the .90s) and is regarded as a moderately valid predictor of law school grades. In one fascinating study, LSAT scores correlated more strongly with state bar examination results than with law school grades (Melton, 1985).

The MCAT is required of applicants to most medical schools in the United States. Unlike the LSAT, which has minimal dependence on specific information, the MCAT assesses knowledge of biology, chemistry, and physics, as well as verbal and quantitative thinking skills. The MCAT yields six scores: biology, chemistry, physics, science problems, skills analysis: reading, and skills analysis: quantitative. Each of the six scores is reported on a scale from 1 to 15 (means of about 8.0 and standard deviations of about 2.5). Reliability of the MCAT is somewhat lower than that of other aptitude tests used for admissions testing, with internal consistency and split-half coefficients mainly in the low .80s. MCAT scores are mildly predictive of performance in medical school. As with many aptitude tests used for selection purposes, the validity of the MCAT is no doubt substantially underestimated by studies that show weak correlations between MCAT scores and performance in medical school. This is due to the restriction-of-range conundrum noted above: examinees with low MCAT scores—who would generally confirm the validity of the test by performing poorly in medical school—are rarely admitted.

TABLE 2
Subtests of the ASVAB

Arithmetic reasoning*	General science
Numerical operations*	Mathematics knowledge
Paragraph comprehension*	Electronics information
Word knowledge*	Mechanical comprehension
Coding speed	Automotive and shop information

*Armed Forces Qualifying Test (AFQT).

MILITARY AND OCCUPATIONAL APPLICATIONS

Another important use of aptitude tests is for vocational guidance and selection in military and occupational settings. The most important test used by military personnel is the ARMED SERVICES VOCATIONAL APTITUDE BATTERY (ASVAB). The ASVAB is a paper-and-pencil group test used by the armed services to screen applicants and to assign recruits to jobs and training programs. The ASVAB consists of ten subtests (Table 2). Four of the subtests constitute the Armed Forces Qualification Test (AFQT), which is the initial qualifying test for all services.

ASVAB results consist of various subtest combinations reported as seven composite scores. The three academic composites are as follows:

- Academic ability
 - Word knowledge
 - Paragraph comprehension
 - Arithmetic reasoning
- Verbal
 - Word knowledge
 - Paragraph comprehension
 - General science
- Mathematics
 - Mathematics knowledge
 - Arithmetic reasoning

The four occupational composites are as follows:

- Mechanical and crafts
 - Arithmetic reasoning
 - Mechanical comprehension
 - Auto and shop information
 - Electronics information

- Business and clerical
 - Word knowledge
 - Paragraph comprehension
 - Mathematics knowledge
 - Coding speed
- Electronics and electrical
 - Arithmetic reasoning
 - Mathematics knowledge
 - Electronics information
 - General science
- Health, social, and technology
 - Word knowledge
 - Paragraph comprehension
 - Arithmetic reasoning
 - Mechanical comprehension

The ASVAB composite scores show excellent reliabilities (internal consistency coefficients typically in the low .90s) and good predictive validities. For example, the median correlation, corrected for restriction of range, between ASVAB composites and final grades in navy entry-level vocational schools was .73 (Weitzman, 1985).

One concern about the ASVAB is the redundancy of the composite scores. The average correlation among these seven scores is .86, which suggests that

the composites do not measure separate, specific aptitudes (Murphy, 1984). The ASVAB is a good measure of general ability, but its functioning as a multiple aptitude test battery is more controversial.

The Differential Aptitude Tests (DAT) are widely used for educational and vocational guidance, particularly for students in grades 8 through 12 (Bennett, Seashore, & Wesman, 1982). In addition, the DAT is used for vocational guidance of young adults and in the selection of employees. The subtests and item types of the DAT are described in Table 3.

DAT scores are reported as percentile ranks in comparison to the performance of 62,000 students from representative high schools in the United States. In addition to scores for the eight tests listed in Table 3, a ninth score indicative of scholastic aptitude is also provided. This score is based upon the average of verbal reasoning and numerical ability.

The individual components of the DAT possess excellent reliability, but the question of validity is more complicated. There is no doubt that DAT scores correlate significantly with relevant nontest criteria, such as course grades and success in vocational training programs. This is especially true for the scholastic-aptitude component (verbal reasoning and numerical ability), which correlates in the .70s and .80s with high

TABLE 3
Subtests and item types on the DAT

<i>Subtest</i>	<i>Item Types</i>
Verbal reasoning (50 items, 30 minutes)	Solve verbal analogies.
Numerical ability (40 items, 30 minutes)	Perform arithmetic computations.
Abstract reasoning (45 items, 20 minutes)	Detect relationships in a series.
Clerical speed and accuracy (2 parts: 100 items, 3 minutes each)	Compare letter and number combinations quickly.
Mechanical reasoning (70 items, 30 minutes)	Apply mechanical principles.
Spatial relations (60 items, 25 minutes)	Visualize in three dimensions.
Spelling (90 items, 20 minutes)	Identify correct and incorrect spelling.
Language usage (50 items, 20 minutes)	Identify mistakes in grammar, punctuation, or capitalization.

school and college grades. Yet, the DAT correlations do not always support the differential validity of the individual tests. For example, the best single predictor of boys' grades in English and literature is numerical ability, not verbal reasoning. In spite of such cautionary findings, DAT results can be used to suggest specific educational and occupational areas related to each of the aptitude scores. For example, persons who score high on spatial relations might find architecture a suitable profession; a high score on clerical speed and accuracy is compatible with the office professions; a high score on mechanical reasoning might signify aptitude for a hands-on profession such as mechanic. These kinds of recommendations, which are included in ancillary materials provided with the DAT results, should be viewed as discussion points for students and their parents, and not regarded as firm predictions.

TESTS FOR SPECIAL APTITUDES

Hundreds of narrow aptitude tests have been constructed to meet specialized needs. These tests help perform the same kinds of gatekeeping and guidance functions as the multiple aptitude batteries discussed above. For example, in selecting students for admission, an art school might use an artistic aptitude test in conjunction with other information; in deciding whether to pursue intensive training in music, a student might desire feedback about her aptitude in that area; and in admitting young students to a program for the gifted and talented, a psychologist might rely, in part, on a creativity test. There are many situations in which short tests for special aptitudes are essential. For illustrative purposes, two special aptitude tests will be described here, the Bennett Mechanical Comprehension Test, or BMCT, and the Seashore Measures of Musical Talent, or SMMT. (For further coverage of this topic, the reader is referred to the *Mental Measurements Yearbook*, published periodically by the Buros Institute of Mental Measurements; e.g., Kramer & Conoley, 1992).

The BMCT can be used to help select applicants for training in hands-on professions that require the understanding of mechanical principles (e.g., plumber, machinist, or mechanic). The test consists of pictures that depict basic mechanical principles encountered in everyday life. For example, a typical item might show

the drawing of a bus and ask the examinee to designate the seat that provides the smoothest ride. A huge body of research has been published on the BMCT (Bechtoldt, 1972). Interestingly, this test proved to be one of the best predictors of pilot success during World War II (Ghiselli, 1966).

The SMMT has been widely used to help identify persons with aptitude in music. The six tests of the SMMT measure pitch, loudness, time, rhythm, timbre discrimination, and tonal memory. The format of the test is deceptively simple: after appropriate instructions and sample items, pairs of musical stimuli are presented, and for each pair, the examinee must determine whether the stimuli are the same or different on the dimension in question. Reviewers generally respect the SMMT, even though they complain that it overlooks important aspects of musical talent (Bean, 1965).

APTITUDE TESTS IN PERSPECTIVE

In modern industrialized societies, aptitude tests have been widely accepted as proper starting points for the selection of applicants in many educational, employment, and military settings. These tests are also generally respected as useful tools in academic counseling, career advising, and job placement, particularly when the results are conveyed and interpreted by a trained counselor or psychologist. These positive features notwithstanding, there is room for improvement in aptitude testing, and certain cautions and concerns about these instruments need to be mentioned.

Regarding the improvement of aptitude tests, one very promising development is computerized adaptive testing (CAT). The essential feature of this approach is flexible, individualized administration of test items by a computer. In CAT, a computer program monitors the examinee's performance item by item and selects appropriately difficult questions, based on results up to that point. Each subject is administered the minimum number of questions needed to obtain a predetermined level of measurement accuracy. The subject's score is based not on the number of items passed but on their difficulty level. Even though each examinee may answer a different set of questions, the scores are comparable for all persons examined with the item pool (Anastasi, 1988). As computer systems continue to be-

come more accessible and interchangeable, CAT and other computerized approaches to testing likely will flourish.

The use of aptitude tests should be tempered with practical cautions and philosophical concerns. The most important practical caution is that aptitude tests are group tests and therefore vulnerable to the well-known pitfalls of such instruments. In group testing, it is difficult to know whether the examinee has followed directions correctly and put forth a good effort. Likewise, it is difficult to determine whether the examinee has a specific learning disability (e.g., a reading disability) as opposed to low aptitude. Invariably, some examinees will score far below their true ability on aptitude tests, with undesirable consequences for the individual and society as well. Aptitude-test users need to acknowledge the fallibility of their instruments and incorporate checks and balances in decision making, where possible.

The use of aptitude tests also raises philosophical concerns about the heavy emphasis upon objective test scores for selection purposes such as college admission. E. Kifer (1985) raised this point in reference to the ACT: "It is just as defensible to select on talent broadly construed as it is to use test scores however high. There are talented students in many areas—leaders, organizers, doers, musicians, athletes, science award winners, opera buffs—who may have moderate or low ACT scores but whose presence on a campus would change it." The proper basis for selection is, of course, a question of societal values and need not involve aptitude tests at all. Nonetheless, to the extent that society can identify specific human capacities desired for selection, an appropriate aptitude test is probably the best way to identify persons with those capacities.

(See also: ARMED SERVICES VOCATIONAL APTITUDE BATTERY; GENERAL APTITUDE TEST BATTERY; TESTING IN GOVERNMENT AND INDUSTRY.)

BIBLIOGRAPHY

- AMERICAN COLLEGE TESTING PROGRAM. (1988). *Technical manual for the ACT assessment*. Iowa City, IA: Author.
- ANASTASI, A. (1988). *Psychological testing*. New York: Macmillan.
- BEAN, K. L. (1965). Review of the Seashore Measures of Musical Talents. In O. K. Buros (Ed.), *The sixth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- BECHTOLDT, H. P. (1972). Review of the Bennett Mechanical Comprehension Test. In O. K. Buros (Ed.), *The seventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- BENNETT, G. K., SEASHORE, H. G., & WESMAN, A. G. (1982). *Differential Aptitude Tests: Administrator's handbook*. San Antonio, TX: The Psychological Corporation.
- COHN, S. J. (1985). Review of the GRE. In J. V. Mitchell, Jr. (Ed.), *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- DONLON, T. F. (ED.). (1984). *The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board.
- GHISELLI, E. E. (1966). *The validity of occupational aptitude tests*. New York: Wiley.
- GREGORY, R. J. (1992). *Psychological testing: History, principles, and applications*. Boston: Allyn & Bacon.
- HAMMILL, D. D. (1985). *Detroit Tests of Learning Aptitude (DTLA-2)*. Austin, TX: PRO-ED.
- JACKSON, D. N. (1984). *Manual for the Multidimensional Aptitude Battery*. Port Huron, MI: Research Psychologists Press.
- JAEGER, R. M. (1985). Review of the GRE. In J. V. Mitchell, Jr. (Ed.), *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- KIFER, E. (1985). Review of ACT Assessment Program. In J. V. Mitchell, Jr. (Ed.), *The ninth mental measurements yearbook*. Lincoln: University of Nebraska Press.
- KRAMER, J. J., & CONOLEY, J. C. (EDS.). (1992). *The eleventh mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- MELTON, G. B. (1985). Review of Law School Admissions Test. In J. V. Mitchell, Jr. (Ed.), *The ninth mental measurements yearbook*. Lincoln, NE: Buros Institute of Mental Measurements.
- MURPHY, K. R. (1984). Armed Services Vocational Aptitude Battery. In D. J. Keyser & R. C. Sweetland (Eds.), *Test critiques* (Vol. 1). Kansas City, MO: Test Corporation of America.
- WECHSLER, D. (1981). *Manual for the Wechsler Adult Intelligence Scale-Revised*. San Antonio, TX: The Psychological Corporation.

WEITZMAN, R. A. (1985). Review of Armed Services Vocational Aptitude Battery. In J. V. Mitchell, Jr. (Ed.), *The ninth mental measurements yearbook*. Lincoln, NE: Bureau of Mental Measurements.

ROBERT J. GREGORY

APTITUDES See ABILITIES AND APTITUDES.

APTITUDE-TREATMENT INTERACTION

The term *aptitude-treatment interaction* (ATI) refers to situations in which particular characteristics of different persons influence the effects of treatment conditions administered to those persons. In turn, the importance of particular personal characteristics in relation to the goals of treatment depends on what particular treatment conditions are administered. In other words, individual differences among persons (aptitude variables) interact with variations in environmental conditions (treatment variables) to influence learning, job performance, or some other important goal; the aptitude and treatment effects are multiplicative rather than additive; that is, they jointly determine a person's behavior.

Possible aptitude variables include not only intelligence and various specialized cognitive or psychomotor abilities but also many aspects of human personality and motivation; only intellectual aptitudes are considered here. Treatment variables may be alternative instructional methods or media in educational settings, alternative training systems in industry and the military, or alternative psychotherapeutic techniques in clinical settings. Treatment variations may also include job or equipment design alternatives where optimal work performance, not just learning, is the goal; classifications of persons into different jobs are basically aptitude-treatment matching decisions implying interactions.

The study of ATI in education and training aims at understanding when, how, and why different kinds of persons benefit from different kinds of instruction, in the hope that instructional treatment conditions can be improved by adapting them to the learning needs and characteristics of each kind of person. Similarly,

the therapist seeks to choose or create a therapy adapted to the particular needs and characteristics of each client; research compares the effectiveness of different therapies against the initial characteristics of different clients. Job or equipment design issues usually involve the study of interactions to evaluate personnel selection, classification, and training procedures or to suggest workplace redesigns.

Research on ATI is a special case of the scientific study of organism-environment interaction. The possibility of such interactions has long been routinely acknowledged in the scientist's standard qualifier "other things being equal" and the routine question "Can we generalize to other groups (communities, cultures, etc.)?" Evolutionary biology is interactionist. ATI is put to practical use in medicine; for example, the physician's choice of antibiotic treatment depends on the patient's answer to the test question "Have you ever had an allergic reaction to penicillin?" This example also shows that aptitude can be positive or negative.

Although interactional thinking is evident in the writings of ancient philosophies, both Eastern and Western, it was Lee J. CRONBACH (1957) who defined ATI as both a fundamental concept and a fundamental problem for modern psychology. The scientific and practical implications of the problem have been systematically addressed only since the 1970s (Cronbach, 1975; Cronbach & Snow, 1977; Snow, 1989), but interactional research has now accumulated in a variety of fields of psychology. There have also been substantial advances in methodology (Cronbach, 1982, 1991; Snow, 1991a).

EVIDENCE AND IMPLICATIONS FOR EDUCATION

The most extensive evidence on ATI so far comes from research in education. Many kinds of individual differences among students have been observed and measured. When these measures predict student differences in learning from instruction, they are considered indicators of aptitude, which should be interpreted as readiness to profit from the particular instructional treatment at hand. Indeed, intelligence tests were first invented to predict school achievement under the conventional instructional conditions in Paris around

1900. Unfortunately, because these predictive correlations have since been found so strong in so many educational situations, intelligence came to be equated with aptitude and interpreted as the ability to learn from any kind of instruction. But measures of intelligence do not provide uniformly high correlations; they correlate more highly with learning when instruction is incomplete, complex, and relatively unstructured, and less highly as instruction is more complete, carefully structured, and controlled by teachers. This is the most strongly supported ATI finding. Also, other measures not reflecting conventional differences in intelligence correlate with learning outcomes and thus indicate other sources of aptitude. Special abilities, previous school achievements, and various cognitive- and learning-style differences are examples of cognitive aptitudes beyond those represented by conventional intelligence tests. As noted earlier, personality differences may also indicate aptitude differences for certain kinds of instruction and may combine with intellectual differences in doing so. When any aptitude measures show differential predictive relations to learning under different instructional conditions, this signifies ATI. The problem then is to discover what aspects of the aptitude differences interact with what aspects of the instructional conditions and why they do so.

Many kinds of aptitude differences among students and many kinds of instructional treatments have now been studied. Since ATI findings often occur in education, there is no doubt that ATI exists, but for most ATI hypotheses, evidence is spotty. Even for the strong ATI finding with general-intelligence measures noted above, theoretical and practical understanding of ATI is not yet sufficient to allow its routine use in educational planning or instructional design. Nonetheless, it is clear that routine use of ATI methodology is both possible and necessary in educational-program evaluations. Any evaluation aimed at comparing alternative teaching methods or environments must ask not only which treatment is best on average but also which treatment is best for each of the individual learners to be served.

ATI findings are of theoretical interest because they demonstrate construct validity for aptitude and learning measures in a new way: they show how aptitude-

learning relations can be experimentally manipulated and thus understood in a causal rather than merely a correlational framework. This suggests that neither aptitude constructs nor educational learning processes can be fully understood without reference to one another and raises the important possibility that common psychological processes and structures underlie both aptitude and learning differences. Furthermore, as ATI appears ubiquitous in education, it is necessary to understand not only persons and situations as independent foci but also the reciprocal interrelations that define the interface between them. Learning and thus aptitude come to be seen as situated in the person-situation union, not just in the heads of persons. This is a new theoretical and philosophical problem for psychology and education (Snow, 1992).

Practical interest stems from the possibility that ATI can be used to adapt teaching to fit different learners optimally. Many attempts at individualizing instruction have failed to eliminate individual differences in learning outcomes because they adapted only to limited aspects of student performance, for example, by allowing differences in pace. The hope is that research on ATI can provide decision rules that indicate how to vary instructional conditions in other ways that mesh with particular learner strengths while avoiding particular learner weaknesses. A related hope is that such research will indicate how best to develop aptitudes directly for persons with different initial aptitude profiles. Again, the general question for research, as well as for evaluation studies, is which of the available or conceivable teaching methods, media, or environments is most likely to provide equality of educational opportunity to each individual learner, for aptitude development and for educational achievement, despite the diversity of initial aptitude profiles in any group of persons to be served. The commitment to optimal diversity of educational opportunity also demands that educational environments be chosen or invented and evaluated using a perspective that includes ATI (Corno & Snow, 1986).

However, ATI research has shown that interactions are often complex and difficult to pin down. No simple or general principles for matching students and teachers, teaching methods, or school environments have emerged. This is due in part to the difficulties involved

in conducting ATI research and to the widespread lack of understanding of appropriate methods. But it is due also to the multidimensional, dynamic, often local and even transient character of the person-situation interface. Results to date suggest that work toward instructional theories that seek to optimize instruction for individuals in real school settings will need to be built up from continuous local diagnosis, description, and evaluation activities; local instructional models rather than general prescriptive theories seem to be the more realizable goals in education (Snow, 1991b).

AN EXAMPLE OF ADAPTIVE INSTRUCTION

An adaptive instructional system such as that envisioned by R. Glaser (1977; cf. Corno & Snow, 1986) could be based on ATI. For example, it might offer two or three instructional treatments representing alternative routes to successful attainment of some defined achievement level in a course or series of courses. An initial diagnosis of aptitude for learning in each available treatment would suggest which route should be taken by each learner. Also available in such a system would be at least one form of compensatory, direct training of aptitude for learners diagnosed as unready for any of the available instructional treatments.

Two alternative instructional treatments can be suggested for inclusion in such a system, since they have been frequently evaluated using measures of general intelligence, prior scholastic ability, or generalized achievement as aptitude. From that research (Snow, 1982, 1989), it appears that instructional treatments differ in the information-processing burdens they place on, or remove from, the responsibility of the learner, and as a result, the relations of achievement outcome to prior ability become stronger or weaker, respectively. As the treatment requires students to infer missing elements, puzzle things out for themselves, structure and organize their own study, and build their own comprehension, more-able learners do well—they can capitalize on their strengths profitably—while less-able learners do poorly. Yet, the more the instructional treatment relieves students of difficult reading, independent reasoning, and analysis of com-

plex concepts by imposing teacher structure, control, and simplification, the more such treatments seem to circumvent the less-able learners' weaknesses. Unfortunately, the structured and simplified treatments that seem to help less able learners are not optimal for able learners, relative to the more complex and less structured treatments where they excel.

Thus, one alternative treatment might be designed to provide relatively unstructured and minimal guidance and to encourage learner self-direction in a discovery-oriented approach. The teacher might guide the inductive process, but instruction would clearly be student-centered. In contrast, a second treatment might be designed to break down the learning task to give clear step-by-step guidance, feedback, and correction through a series of small units, with frequent summary and review, and simplified demonstrations of the concepts to be learned. Students would be assigned to one or the other treatment on the basis of prior scholastic ability scores taken at the start of instruction. Periodic aptitude and achievement assessments would show the degree to which outcome criterion levels were being achieved for each learner in the particular treatment assigned.

For those students who might not be expected to profit from either alternative initially, compensatory aptitude training would be assigned. This might consist of directed work on academic learning, reasoning, and reading skills, study habits, self-monitoring, and related self-management skills. The aim of such training would be to develop readiness for the structured treatment. Continuing aptitude training would also seek to develop the more independent style required by the unstructured treatment. The aim of continuing assessment would be to shift students among treatments to optimize outcome. But continuing ATI evaluation would also bring in other aptitudes, as prior evidence might suggest. In this example, student anxiety might be considered an additional aptitude because it has often been found that more anxious students do relatively poorly under unstructured or student-centered forms of instruction, as compared with teacher-structured conditions, whereas less anxious students often do not need or want teacher structure. Ability and anxiety appear to combine in higher-order interaction in relation to this treatment contrast.

This example offers a hypothetical adaptation based on ATI evidence, but no such system can be properly designed in the abstract. Local instructional and populational conditions must be considered, aptitude measures and alternative treatment designs must be adjusted to these conditions, and periodic local evaluation is needed.

PROSPECTS FOR RESEARCH AND DEVELOPMENT

It is noteworthy that much research in instructional psychology contrasts the two treatment examples above without evaluating ATI. Much past research on teaching has compared direct teacher-centered instruction with guided discovery. New research on instructional technology considers essentially the same contrast between mastery-oriented computerized tutoring and discovery learning in computerized micro-worlds (Glaser & Bassok, 1989). Also, a large number of cognitive-style and learning-style hypotheses have been developed for use by teachers in adapting classroom instruction to student differences; each such style construct is an ATI hypothesis, though none have yet been adequately evaluated as such. Thus, one important prospect for the future is to conduct evaluative research on these new developments from an ATI perspective.

A second important line for further research aims at improved analysis of the many different kinds of aptitude constructs and measures that have been used in ATI work. New factor-analytic research has clarified the structure of interrelations among broad and narrow cognitive abilities and has shown how best to represent these constructs in ATI studies (Carroll, 1993; Gustafsson, 1989). Furthermore, several lines of cognitive task analysis applied to these ability tests and constructs have identified underlying component processes and strategies on which individuals may differ (Lohman, 1989; Snow, 1989; Sternberg, 1985a, 1985b). Still another research approach traces the individual differences in learning activities engaged in during instruction that appear to mediate aptitude-achievement relations. Task analyses of instructional conditions also suggest mediational differences among alternative treatments that may control aptitude-achievement relations. Important new work is inves-

tigating prior knowledge differences as cognitive aptitudes for instruction (Dochy, 1992; Schneider & Weinert, 1991).

Finally, more intensive analysis of achievement measures may help diagnose in detail the particular kinds of cognitive effects that derive from particular aptitude-treatment combinations. Future work on all these fronts can be brought together within the ATI framework.

BIBLIOGRAPHY

- CARROLL, J. B. (1993). *Human cognitive abilities*. New York: Cambridge University Press.
- CORNO, L., & SNOW, R. E. (1986). Adapting teaching to individual differences among learners. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 605-629). New York: Macmillan.
- CRONBACH, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, *12*, 671-684.
- CRONBACH, L. J. (1975). Beyond the two disciplines of scientific psychology. *American Psychologist*, *30*, 116-127.
- CRONBACH, L. J. (1982). *Designing evaluations of educational and social programs*. San Francisco: Jossey-Bass.
- CRONBACH, L. J. (1991). Emerging views on methodology. In T. D. Wachs & R. Plomin (Eds.), *Conceptualization and measurement of organism-environment interaction* (pp. 87-104). Washington, DC: American Psychological Association.
- CRONBACH, L. J., & SNOW, R. E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.
- DANCE, K. A., & NEUFELD, R. W. J. (1988). Aptitude-treatment interaction research in the clinical setting: A review of attempts to dispel the "patient uniformity" myth. *Psychological Bulletin*, *104*, 192-213.
- DOCHY, F. J. R. C. (1992). *Assessment of prior knowledge as a determinant for future learning*. Utrecht, Netherlands: Uitgeverij Lemma B.V.
- ENDLER, N. S., & MAGNUSSON, D. (EDS.). (1976). *Interactional psychology and personality*. Washington, DC: Hemisphere.
- GLASER, R. (1977). *Adaptive education: Individual diversity and learning*. New York: Holt, Rinehart & Winston.
- GLASER, R., & BASSOK, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology*, *40*, 631-666.
- GUSTAFSSON, J. E. (1989). Broad and narrow abilities in re-

- search on learning and instruction. In P. Kanfer, P. L. Ackerman, & R. Cudeck (Eds.), *Abilities, motivation, and methodology* (pp. 203–237). Hillsdale, NJ: Erlbaum.
- LOHMAN, D. F. (1989). Human intelligence: An introduction to advances in theory and research. *Review of Educational Research*, 59, 333–373.
- MAGNUSSON, D., & ALLEN, V. L. (EDS.). (1983). *Human development: An interactional perspective*, New York: Academic Press.
- PERVIN, L. A., & LEWIS, M. (1978). *Perspectives in interactional psychology*. New York: Plenum.
- SCHNEIDER, W., & WEINERT, F. E. (EDS.). (1990). *Interactions among aptitudes, strategies, and knowledge in cognitive performance*, New York: Springer-Verlag.
- SNOW, R. E. (1982). Education and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 493–585) New York: Cambridge University Press.
- SNOW, R. E. (1989). Aptitude–treatment interaction as a framework of research in individual differences in learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 13–59). New York: W. H. Freeman.
- SNOW, R. E. (1991a). Aptitude–treatment interaction as a framework for research on individual differences in psychotherapy. *Journal of Consulting and Clinical Psychology*, 59, 205–216.
- SNOW, R. E. (1991b). The concept of aptitude. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in social science* (pp. 249–284). Hillsdale, NJ: Erlbaum.
- SNOW, R. E. (1992). Aptitude theory: Yesterday, today and tomorrow. *Educational Psychologist*, 27, 5–32.
- SNOW, R. E., & LOHMAN, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology*, 76, 347–376.
- SNOW, R. E., & LOHMAN, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. L. Linn (Ed.), *Educational Measurement* (3rd ed., pp. 263–331). New York: Macmillan.
- SNOW, R. E., & SWANSON, J. (1992). Instructional psychology: Aptitude, adaptation, and assessment. *Annual Review of Psychology*, 43, 583–626.
- STERNBERG, R. J. (1985a). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge: Cambridge University Press.
- STERNBERG, R. J. (ED.). (1985b). *Human abilities: An information processing approach*. San Francisco: W. H. Freeman.

RICHARD E. SNOW

ARITHMETIC *See* WAIS-R SUBTESTS.

ARMED SERVICES VOCATIONAL APTITUDE BATTERY When the United States entered World War I, the need for group intelligence tests to select and classify military applicants became readily apparent. The ARMY ALPHA AND BETA TESTS were subsequently developed and administered to nearly 2 million men—an event that helped firmly establish intelligence testing in the United States. Since that era, numerous changes in military tests have occurred, leading ultimately to the current test battery known as the Armed Services Vocational Aptitude Battery (ASVAB).

The ASVAB was introduced in 1976 to achieve greater consolidation of testing programs across military services. For example, between World War II and the Vietnam War, each U.S. military service maintained its own independent testing program. Although for many of those years all services employed the same tests for enlistment decisions, service-specific tests were used to assign the enlistees to military jobs. These independent testing programs ended with the introduction of the ASVAB. Since 1976, the ASVAB has been used by all military services both to determine mental qualifications for enlistment and eligibility for specific job assignments. The specific reasons for adopting a joint-service battery were to (1) avoid subjecting multiple-service applicants to more than one test session; (2) facilitate interservice referrals of applicants; and (3) enable service psychologists to focus their efforts on a single enlisted selection and classification battery (Foley & Rucker, 1989).

Approximately 2 million examinees per year are currently administered the ASVAB. Examinees include applicants for military service in addition to high school students who take the ASVAB for vocational counseling purposes. Vocational counseling is feasible because, through the use of validity generalization techniques and linkages between the ASVAB and civilian batteries, such as the General Aptitude Test Battery and the Differential Aptitude Tests, predictions for military validity studies can be extended to much of the civilian occupational spectrum. New forms of the ASVAB are produced at regular intervals and care-

fully equated to ensure that the test remains useful for its various purposes. As of 1993, alternate forms of the ASVAB were being developed by the Defense Management Data Center in Monterey, California.

ASVAB CONTENT

The ASVAB was designed to be a differential aptitude battery, meaning that in theory the specific tests should, relative to one another, measure clearly defined, nonoverlapping aptitudes and be predictive of performance in different jobs or job families. The ASVAB consists of the ten independently timed and

scored subtests shown in Table 1, all of which are power tests with the exception of Numerical Operations and Coding Speed, which are speeded. Power tests consist of items of varying difficulty, and are designed to measure the absolute level of performance that an individual can attain under relatively generous time limits. By contrast, speeded items are easy for all examinees to answer correctly if given sufficient time.

Factor analyses of ASVAB scores typically indicate that a four-factor solution is optimal (e.g., Bock & Moore, 1984). In other words, the ten ASVAB subtests are interrelated in such a way that they actually measure four, rather than ten, ability dimensions. The four

TABLE 1
ASVAB subtests

<i>Subtest</i>	<i>Abbreviation</i>	<i>Description</i>
General Science	GS	A 25-item test of general knowledge of the physical (13 items) and biological (12 items) sciences, administered with an 11-minute time limit
Arithmetic Reasoning	AR	A 30-item test of ability to solve arithmetic word problems, which typically involve simple calculation based on recognition and application of four basic arithmetic operations (addition, subtraction, multiplication, and division), administered with a 36-minute time limit
Word Knowledge	WK	A 35-item test of knowledge of vocabulary, including words embedded in sentences (11 items) and synonyms (24 items), administered with an 11-minute time limit
Paragraph Comprehension	PC	A 15-item test of reading comprehension based on questions about short paragraphs, administered with a 13-minute time limit
Numerical Operations	NO	A 50-item speeded test of ability to quickly add, subtract, multiply, and divide one- and two-digit numbers, administered with a 3-minute time limit
Coding Speed	CS	An 84-item speeded test of ability to substitute numeric codes for verbal material, administered with a 7-minute time limit
Auto and Shop Information	AS	A 25-item test of everyday knowledge of automobiles, shop practices, and use of tools, administered with an 11-minute time limit
Mathematics Knowledge	MK	A 25-item test of knowledge of high-school-level mathematics (algebra, geometry, fractions, decimals, and exponents), administered with a 24-minute time limit
Mechanical Comprehension	MC	A 25-item test of knowledge of mechanical and physical principles, containing several pictorial items, administered with a 19-minute time limit
Electronics Information	EI	A 20-item test of knowledge of electronics, radio, and electrical principles and information, administered with a 9-minute time limit

NOTE: Scores are reported as Standard Scores, having a mean of 50 and a standard deviation of 10 for a representative sample of 19- to 23-year-old U.S. youth.

dimensions and the subtests included in each dimension are (1) Verbal Ability, measured by the Word Knowledge (WK), Paragraph Comprehension (PC), and General Science (GS) subtests; (2) Clerical Speed, measured by the Numerical Operations (NO) and Coding Speed (CS) subtests; (3) Mathematical Ability, measured by the Arithmetic Reasoning (AR) and Mathematical Knowledge (MK) subtests; and (4) Technical Knowledge, measured by the Mechanical Comprehension (MC), Electronics Information (EI), and Auto and Shop Information (AS) subtests. Moreover, the four dimensions are themselves substantially inter-correlated, suggesting that the battery measures a general intelligence factor as well as the four factors cited above.

Enlistment eligibility is determined for all services by a test composite known as the Armed Forces Qualification Test (AFQT). The AFQT is calculated in several steps. First, the standardized, double-weighted sum of a verbal composite (VE) (a combination of the WK and PC subtests) is added to the standardized scores for the AR and MK subtests. Then the resulting sum is converted to a cumulative percentile score, referred to as the AFQT, normed to the 1980 U.S. youth population (Bock & Moore, 1984). For reporting purposes, the percentile AFQT scores are grouped into six broad categories as shown in Table 2.

Minimum qualifying AFQT percentile scores are regularly adjusted in response to the changing quality and quantity of personnel needed by the armed forces, as well as changes in the status of the recruiting market (Foley & Rucker, 1989). In addition, "cutting scores" can be influenced directly by the actions of

Congress (Eitelberg et al., 1984). For example, Congress has mandated that no applicant with a score below the 10th percentile (i.e., no Category V applicants) may be enlisted, because of the high probability of ineffective performance.

Numerous combinations of ASVAB subtests are used for assigning personnel to the wide range of military jobs. Examples are the Electronics Composite used by all services (derived by combining the AR, MK, GS, and EI scores) and the Clerical Composite used by three services (derived by combining the VE, NO, and CS scores). The remaining automated "assignment formulas" are much too numerous to describe here. Of note, however, is the fact that they are empirically derived and continually updated so that assignment for any particular job is always based on current knowledge about which test combination is most predictive of job performance.

RELIABILITY AND VALIDITY

Retest reliabilities (after an interval of five weeks) for the ASVAB subtests range from the mid-.70s to the mid-.80s (Moreno & Segall, 1992). The sole exception is the PC subtest, which appears to have a reliability of slightly less than .50. With the possible exception of PC, the reliabilities are quite adequate, since ASVAB subtests are always used in composites, which have higher reliabilities than those of single tests.

The validity of the ASVAB has been the subject of much research. A number of recent studies suggest that ASVAB is a good predictor of performance both for students in military training schools and for job incumbents. Ree and Earles (1992), for example, report mean correlations in the .50s and .60s between individual ASVAB subtests and final school grades in numerous Air Force training courses; correlations of grades with ASVAB composites were in the .60s to .70s. Similarly high correlations of ASVAB scores with job performance factors, such as technical proficiency, have also been reported (McHenry et al., 1990), although other studies (e.g., Ree, Earles, & Teachout, 1992) suggest more modest, though still highly significant, relationships. (It should be noted that these are not "raw" correlations. Rather, they have been corrected for the range restriction that results whenever

TABLE 2
Mental ability categories based on AFQT scores

<i>Category</i>	<i>AFQT Percentile Score Range</i>
I	93 and above
II	65-92
IIIa	50-64
IIIb	31-49
IV	10-30
V	1-9

test scores are actually used to disqualify low-scoring applicants). In general, the correlations of ASVAB scores with complex or "information-rich" tasks (e.g., those involving high technology) are considerably (and understandably) higher than the correlations of ASVAB with intellectually simple tasks, such as those emphasizing manual labor.

Regarding ASVAB validities, some researchers (e.g., Hunter, 1986; Ree & Earles, 1992) have argued that the ASVAB is not truly functioning as a differential aptitude battery because in predicting diverse military performance criteria, an ASVAB general intelligence score (psychometric g -score) is almost always more valid than any other test score or score combination that can be derived from the battery. This, of course, suggests that the individual tests as used today do not provide the sort of nonoverlapping information that was apparently the goal of the battery's designers.

COMPUTERIZED ASVAB

For much of the 1980s, research was conducted on a Computerized Adaptive Testing (CAT) version of the ASVAB, the CAT-ASVAB. In adaptive tests, test difficulty is continuously readjusted as a function of the examinee's ongoing performance. Therefore, an examinee who has correctly answered an item will receive a harder item, while an examinee who has missed an item will receive an easier question. The CAT-ASVAB system incorporates so-called item response theory as a basis for selecting items. By early 1992, CAT-ASVAB was being used in a limited number of testing sites, marking a historically important transition for military testing and perhaps for testing in general. The reasons for adopting the CAT-ASVAB were to (1) shorten overall testing time (adaptive tests require roughly one-half the items of standard tests); (2) increase test security by eliminating the possibility that test booklets could be stolen; (3) increase test precision at the upper and lower ability extremes; (4) provide a means for immediate feedback on test scores, since the computers used for testing can immediately score the tests and output the results; and (5) provide a means for flexible test start times (unlike group-administered paper-and-pencil tests, where everyone must start and stop at the same time, com-

puter-based testing can be tailored to the examinees' personal schedules).

Should CAT-ASVAB be implemented on a nationwide basis, the content of the battery is likely to be expanded to include tests that exploit the unique capability of computers to measure new aspects of performance, such as response latencies, and to display previously impossible item types, such as visuospatial tests involving objects in motion. By the early 1990s, with such future changes in mind, many military psychologists were engaged in developing new computer-based tests.

BIBLIOGRAPHY

- BOCK, R. D., & MOORE, E. J. G. (1984). *Profile of American youth: Demographic influences on ASVAB test performance*. Washington, DC: Office of the Assistant Secretary of Defense.
- EITELBERG, M. J., LAURENCE, J. H., WATERS, B. K., & PERELMAN, L. S. (1984). *Screening for service: Aptitude and education criteria for military entry*. Washington, DC: Office of the Assistant Secretary of Defense.
- FOLEY, P. P., & RUCKER, L. S. (1989). An overview of the Armed Services Vocational Aptitude Battery (ASVAB). In R. F. Dillon & J. W. Pellegrino (Eds.), *Testing: Theoretical and applied perspectives*. New York: Praeger.
- HUNTER, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340-362.
- MCHENRY, J. J., HOUGH, L. M., TOQUAM, J. L., HANSON, M. A., & ASHWORTH, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, 43, 335-354.
- MORENO, K. E., & SEGALL, D. O. (1992, October). *CAT-ASVAB precision*. Paper presented at the 34th annual conference of the Military Testing Association, San Diego, CA.
- REE, M. J., & EARLES, J. A. (1992). *Subtest and composite validity of ASVAB Forms 11, 12, and 13 for technical training courses (AL-TR-1991-0107)*. Brooks AFB, TX: Armstrong Laboratory.
- REE, M. J., EARLES, J. A., & TEACHOUT, M. S. (1992). *General cognitive ability predicts job performance (AL-TP-1991-0057)*. Brooks AFB, TX: Armstrong Laboratory.

GERALD E. LARSON

ARMY ALPHA AND BETA TESTS OF INTELLIGENCE The development of the Army Alpha test and its companion test, the Army Beta test, was a milestone in U.S. psychology. They have been referred to as one of psychology's most influential contributions to American society, and the testing movement that it helped to spawn dominated psychology and education in the United States for decades.

The United States entered World War I on April 6, 1917. The next day, Robert M. YERKES, the president of the American Psychological Association (APA), dispatched a letter to members of the Executive Council of the association, in which he wrote:

In the present perilous situation it is obviously desirable that the psychologists of the country act unitedly in the interests of defense. Our knowledge and methods are of importance to the military service of our country, and it is our duty to cooperate to the fullest extent and immediately toward the increased efficiency of our Army and Navy. (Yerkes, 1921, pp. 7-8)

The council quickly approved, and the prestigious National Research Council (NRC) formed a Psychology Committee, with Yerkes as chairman. (Unlike other disciplines, such as physics and chemistry, psychology had no formal status with the NRC prior to this time.) In turn, the APA formed twelve committees to deal with the various means by which psychology could contribute. The most famous of these was the Committee on the Psychological Examination of Recruits, also under the chairmanship of Robert Yerkes. Yerkes, later commissioned a major in the Sanitary Corps of the Army Medical Department, immediately assembled a group of well-known psychologists to design a test of intelligence. The group included Walter V. Bingham, Herbert H. Goddard, Thomas H. Haines, Lewis M. TERMAN, F. Lyman Wells, and Guy M. Whipple. Walter Dill Scott was an original member, but during the first meeting he resigned in a dispute with Yerkes over how the committee should proceed. As a civilian employee of the Adjutant General's Office in the War Department—the Army's personnel department—Scott did provide valuable assistance to the project as well as promoting the use of other forms of psychological testing, primarily for occupational placement. He was later awarded the Distinguished Service

Medal by the War Department—the only psychologist so honored.

According to Yerkes (1919), intelligence test scores of army personnel were needed

1. For the discovery of men whose superior intelligence warranted their consideration for promotion, special training or assignment to positions of unusual responsibility or difficulty.
2. For assistance in selecting suitable candidates for officers' training schools, non-commissioned officers' training schools and other special training organizations.
3. For the guidance of personnel adjutants in the assignment of recruits so that organizations might be built in accordance with desirable intelligence specifications or, in the absence of such specifications, so that their different constituent parts, such as the companies of a regiment, should possess approximately the same mental strength, thus avoiding the risk of weak links in the army chain.
4. For the prompt discovery of men whose low-grade intelligence or mental peculiarities rendered them of uncertain value in the army, and the assignment of such individuals to developmental battalions for observation and preliminary training.
5. For the discovery and recommendation or assignment to labor battalions of men obviously so inferior mentally as to be unsuitable for regular military training, yet promising serviceableness in simple manual labor.
6. For the discovery of men whose mental inferiority makes them unfit for any sort of military duty and whose rejection or discharge should therefore be recommended to medical officers. [Note the deference to the Medical Department, an issue of bitter contention even today.]
7. For utilization in connection with the organization of special training groups so that each group might be instructed or drilled in accordance with its mental capacity, thus avoiding the delay incident to dull or awkward individuals and enabling the specially able men to proceed rapidly and ultimately to take special forms of training in preparation for promotion or other forms of responsibility. (pp. 90-91)

Although Yerkes made similar offers of help to both the Army and Navy on behalf of all twelve APA com-

mittees, only the former gave tentative acceptance. The Surgeon General of the Navy refused, although later in the war individual commanders did avail themselves of the services of individual members of the APA committees on an ad hoc basis (McGuire, 1990, pp. 28–29). This negative response was apparently based on the fact that the Navy did not face the same problems as did the Army. It required far fewer personnel and filled its ranks with volunteers rather than depending upon the draft. It was also able to establish higher mental and educational standards and could more easily and quickly discharge unacceptable personnel.

THE ALPHA TEST

For many years, the 1916 Stanford-Binet Test created by Lewis Terman had been the standard by which intelligence was assumed to be measured. It had to be administered by a trained examiner to one subject at a time, however, and it could consume one or more hours. This was obviously not suited to the needs of the military. The committee sifted through dozens of various kinds of test items (questions and/or problems), produced a number of versions of “Test A” (i.e., Test Alpha), and tried them on a number of Army populations. The final version was a paper-and-pencil group test with eight subtests (Yerkes, 1921, pp. 219–234). Included in this format were a number of the now-traditional multiple-choice items. Ten equivalent forms were developed in order to prevent coaching and so that if the security of one form was violated, another could be substituted.

The eight subtests were as follows.

1. *Oral Directions*. The testee was given a series of oral commands, such as, “When I say, ‘Go,’ cross out the letter C [in a series of letters on the test form] and draw a line under the second letter before H. Go!”
2. *Arithmetical Problems*. This subtest included such problems as, “If it takes 6 men 3 days to dig a 180-foot drain, how many men are needed to dig it in half a day?”
3. *Practical Judgment*. This subtest consisted of multiple-choice items, such as, “Why is leather used for shoes? Because: a. it is produced in all countries; b. it wears well; c. it is an animal product.”
4. *Antonyms*. This subtest consisted of pairs of words, such as “wet-dry” and “delicate-tender.” For each pair, the subject was asked to indicate if the two words were the “Same” or “Opposite.”
5. *Disarranged Sentences*. This subtest consisted of sentences in which the words were scrambled, such as, “Lions strong are” and “Months coldest are summer the.” The testee was asked to arrange the words of each sentence in their proper order, and then indicate if the resulting sentence was “True” or “False.”
6. *Number Series Completion*. The subject was shown a series of numbers (e.g., “3, 6, 9, . . .”) and asked to write out the two numbers that should come next.
7. *Analogies*. In each item, the testee was presented with a pair of words and asked to complete a second pair based on the relationship shown in the first pair—for example, “above-below : top-_____ : a. spin; b. bottom; c. surface; d. side.”
8. *Information*. This was another multiple-choice test in which the man completed a sentence by choosing a word that made the sentence truest—for example, “Coral is obtained from: a. mines; b. elephants; c. oysters; d. reefs.”

The men were tested in groups of 50 to 400 under the direction of an Army psychologist. Each subtest was strictly timed. The psychologist would read the direction aloud, and then say, “Go”; when the allotted time had elapsed, he would call out, “Stop.” The entire test took between forty and fifty minutes. The tests were then collected and processed manually by an array of clerks, who would score each test by placing a stencil over the answer sheet—an innovation at the time.

THE BETA TEST

The Beta examination was designed as an alternative to the Alpha for those men who were illiterate and/or foreign born and had little skill with the English language. A man was administered the Beta if he could not comprehend the directions for the Alpha or failed it. A working definition of illiteracy was “someone who could not read a newspaper or write a letter home.”

The Beta contained seven subtests and was also

given as a group test. It consisted primarily of symbols and pictures instead of prose and numbers, although each man was required to write his name, age, and identifying information on the answer sheet (assistance was given if needed).

The directions were given orally by the psychologist, with the assistance of a demonstrator and a large blackboard. The demonstrator would perform a sample problem on the board, such as drawing a line through a maze or counting the number of cubes in a three-dimensional drawing. After each problem was explained and demonstrated, the examinees were given an allotted time in which to solve the problem on their test blanks, after which the examiner would say, "Stop," and proceed to the next problem.

The seven subtests were as follows.

1. *Mazes*. The testee was required to draw a line through a series of mazes without crossing any of the lines. (These are the same problems most schoolchildren find in such publications as comic books or children's magazines.)
2. *Cube Counting*. This subtest contained drawings of piles of blocks, rendered in three dimensions, with the fourth side of the pile not visible. The subject was asked to count the number of blocks in each pile.
3. *X-O Series*. Each item consisted of a line of squares. The first few squares were filled in with a series of Xs and Os in some logical pattern, such as "X,O,X,O,X,O" or "X,X,O,O,X,X,O"; the remaining squares were blank. The testee had to complete the series by using the same sequence to fill in the blanks.
4. *Digit Symbol*. This subtest involved a code in which each of the numbers 1 through 9 was paired with either a letter or some other symbol. The subject was presented with a series of double squares in which the upper square contained a number and the lower square was blank. The task was to consult the code at the top of the page and then as rapidly as possible fill in each blank square with the appropriate number.
5. *Number Checking*. Each item consisted of two numbers, separated by a dotted line—for example, "650 653." Each number contained up to ten digits. The subject was instructed to place an X

on the dotted line if the numbers were identical and leave it blank if they were different.

6. *Pictorial Completion*. The test booklet contained a series of pictures with a part missing, such as a picture of a four-fingered hand and a picture of a fish without an eye. After the task was demonstrated on the blackboard (as in all the subtests), the testees were asked to "fix" the picture by drawing in the missing part.
7. *Geometrical Construction*. The test page contained ten squares. Alongside each square were three geometric shapes—triangles of different shapes, half-moons, and the like. The task was to draw how the three pieces would fit into the larger square by joining the three pieces so that they formed a square within a square.

The men who failed the Beta were then scheduled for individual examinations by the psychologist. They were tested with either an abbreviated version of the Stanford-Binet or the Yerkes-Bridges Point Scale.

Each testee was given an intelligence rating, ranging from "A" ("Very Superior Intelligence; considered High Officer type when combined with other qualities") to "E" ("Mental Inferiority, justifying recommendation for remedial training, rejection, or discharge"). Every man rated "D minus" or higher was deemed to be intellectually suited for military duty, but his subsequent disposition and assignment were determined by the personnel department, which considered his mental level in making the decision.

WARTIME USE

By January 31, 1919, the final versions of the two tests had been used in thirty-five Army cantonments (i.e., bases) and administered to 1,726,966 men. About 0.5 percent were recommended for discharge, while another 3 percent were recommended for placement in labor battalions and/or considered unfit for regular military service (Yerkes, 1921, p. 101). By the end of the war, more than 100 psychologists had been commissioned in the Sanitary Corps of the U.S. Army Medical Department; along with approximately 275 enlisted men, they were given special training in the newly established School of Military Psychology. Many of the enlisted men also had university degrees. (See Yerkes, 1921, pp. 36–38 for a roster of these men.)

METHODOLOGICAL ISSUES

The Alpha and the Beta were found to be statistically reliable in a test-retest situation; that is, they tended to produce the same results each time they were administered. Each test item also correlated significantly with the total score, but not so highly as to suggest it was merely duplication. These qualities are considered requirements of a good psychological test, but the matter of validity was not easily resolved: What did these tests measure—and, most important, did they measure something that was of practical value to the Army?

There is no universally agreed-upon definition of intelligence. Frequently, psychologists will adopt an operational definition that seems suited to a particular application. In describing the use of the tests to commanding officers, Yerkes and his colleagues offered the following rationale:

The tests give a reliable index of a man's ability to learn, to think quickly and accurately, and to comprehend instructions. They do not measure loyalty, bravery, dependability, or the emotional traits that make a man "carry on." A man's value to the service is measured by his intelligence plus other necessary qualifications. (Yerkes, 1921, p. 424)

In a number of studies, both the Alpha (Yerkes, 1921, pp. 327–346) and the Beta (pp. 379–395) correlated significantly with the Stanford-Binet; that is, they tended to measure the same kind of ability. These correlations varied depending upon the group studied, such as "White Adults," "Negro Army recruits," and so forth. While these studies tended to satisfy psychologists that they were measuring intelligence, it did not convince the War Department that these scores in any way related to its needs.

Additional studies were performed in which commanding officers were asked to rate the intelligence and general mental performance of enlisted men and officers with whom they were familiar. Again, the scores on the tests correlated significantly with these ratings, but to a lesser degree than with the Stanford-Binet.

In another demonstration of its applicability to the needs of the Army, an extensive study was undertaken to show the relationship of the two tests to military

rank. It was assumed that, on average, privates would score lower than would, say, sergeants, who in turn would score lower than would commissioned officers. It was also assumed that certain specialists with extensive civilian education, such as physicians, would score higher than would other commissioned officers. To a highly significant degree, this proved to be the case (Yerkes, 1921, pp. 839–860).

These correlations with "the real world of the Army" convinced a number of commanding officers of the basic utility of the tests, and they cooperated in promoting use of the tests, at least within their local commands. (For an 890-page exhaustive description of the entire Army testing program, the ambitious reader is referred to Yerkes, 1921.)

POSTWAR DEVELOPMENTS

In spite of the fact that the Alpha and the Beta were administered to so many thousands of men, the end of the war prevented the tests from being used routinely on an Armywide basis. There was also a great deal of confusion among commanders as to the nature and purpose of the tests. They were sometimes confused with the neuropsychiatric screening program of the Medical Department (which was held in low esteem by some commanders), and some psychiatrists viewed them as infringing upon their prerogatives.

When the war ended, there was less than unanimous acceptance of the testing program, and while work on psychological testing continued within the Office of the Adjutant General (Yerkes's work was almost entirely under the purview of the Medical Department), it was not continued as part of the peacetime Army. Nevertheless, many people believed that the program had increased the visibility and prestige of psychology in the United States, in both industry and academia.

Undaunted by the fact that his enthusiasm was not universally shared, Yerkes declared at the end of the war: "Two years ago, mental engineering [it was Yerkes who coined this term] was a dream of a few visionaries. Today it is a branch of technology which, although created by the war, is evidently to be perpetuated and fostered by education and industry" (Yerkes, 1919, p. 149). He pointed to the development of the Alpha and the Beta as the primary evidence for this conclusion.

Three developments were especially important in the evolution of intelligence testing: (1) the publishing of the 1916 Stanford-Binet; (2) the “invention” of the IQ, by William Stern and Lewis Terman; and (3) the availability of the civilian version of the Army Alpha. These events combined to produce a whole new industry of psychological testing, including tests of personality, attitudes, and interests, and use of the new tests in vocational guidance and employee selection as well as education. Furthermore, this movement dominated psychology in the United States in the decades after World War I.

The multiple-choice format made possible mass testing and the accumulation of huge amounts of data, not only for research but also for public consumption. Between 1920 and World War II, the United States went on what has been called “the IQ binge.” Any person with access to these tests could administer them and use the results as he or she saw fit, in spite of the fact that many users were not qualified to properly administer the tests or apply the results. Despite efforts to restrict use of the tests, they were easily available to school administrators, personnel managers, and college faculty.

The resulting misuse of IQ tests, especially in education, eventually soured the U.S. public, and the field of psychology was burdened for many years with a considerable amount of negative publicity. (For popular accounts of this era see Block & Dworkin, 1976, *The IQ Controversy*, and/or Fancher, 1985, *The Intelligence Men: Makers of the IQ Controversy*.)

BIBLIOGRAPHY

- MCGUIRE, F. L. (1990). Psychology aweigh! A history of clinical psychology in the United States Navy, 1900–1988. Washington, DC: American Psychological Association.
- YERKES, R. M. (1919). Report of the psychology committee of the National Research Council. *The Psychological Review*, 26(2), 83–149.
- YERKES, R. M. (Ed.). (1921). *Psychological examining in the United States Army. Memoirs of the National Academy of Sciences* (Vol. 15). Washington, DC: Government Printing Office.

FREDERICK L. MCGUIRE

ARTIFICIAL INTELLIGENCE *Artificial intelligence* (AI) is the endeavor of building computer-based artifacts that embody the processes of intelligence. Within this definition, however, there are a range of conceptions of the goals, tools, and approaches of AI. For some researchers, AI is a methodology for studying the human mind by building models that instantiate theories of cognition. A working AI program is a demonstration that the principles embodied in the program are sufficient to account for some aspect of human reasoning or behavior. For other researchers, AI is an engineering discipline. Their goal is to build clever computer programs that can perform human-like tasks. For them, human reasoning processes represent little more than a source for ideas on how to solve problems.

In between, of course, there is a large middle ground of researchers who wish to explore the mind through building AI programs, but who believe that the way to do this is to build programs that perform real-world tasks that genuinely need to be done.

Early work in artificial intelligence focused on simple tasks in miniature worlds. It was assumed that the methods used to reason in these “toy domains” would naturally be extensible into real-world programs. However, this has turned out not to be the case. The limited complexity of these domains allowed the use of algorithms which are simply not capable of performing even moderately realistic tasks.

An example of this is seen in planning research, an enterprise established by Allen Newell and Herbert A. SIMON (Newell & Simon, 1963). For example, programs have been written that can plan sequences of actions for moving and stacking blocks. These programs generated all the steps necessary for getting a pile of blocks from one configuration into another. Because of the simplicity of the domain, and the planner’s perfect knowledge of it, it was possible to generate very long lists of moves in advance, before the first piece had even been moved. However, even in this simple universe, this planning strategy is questionable. When the number of blocks in the scenario increases beyond a handful, the program bogs down in an intractably long reasoning process.

Several such lessons have been learned as a result of early AI research. Researchers made a number of assumptions that seemed like plausible guesses about

the nature of intelligence, but experience trying to create intelligent systems has demonstrated that many of them are false. It was only by attempting to build programs that scaled up to complex, real-world tasks that the problems with these assumptions became clear.

Here are three general lessons that have been learned:

- Intelligence consists of numerous mechanisms.

Many of the best-known AI efforts made (and continue to make) the assumption that intelligence is a single, unified process. One of the problems with attacking AI piecemeal, as researchers are forced to do by the scope of the problem, is that they come up with unified architectures that work well for the task they are trying to model. They may assume that this architecture can explain all of cognition, since it works so well for the kinds of cognition they are studying.

There have been major efforts at developing a unified theory of cognition, such as Allan Newell's SOAR program (1990). While it may be possible to build a reasoning system that could reason using one general mechanism for different domains, it appears unlikely that humans reason in this way. Neuroscience and cognitive psychology are providing more and more evidence that the brain consists of many specialized mechanisms for reasoning about different information (such as visual, spatial, verbal, or social phenomena) and for different cognitive functions (such as perceiving the emotions of others, comprehending and generating language, and engaging in intentional metacognitive reasoning). Coming up with computational models that capture the idea of numerous communicating parallel processes is one of the most challenging and exciting areas of AI research.

The idea that intelligence relies on a single mechanism was also implicit in the design of *expert systems*, an application of AI that was widely touted in the 1980s (Feigenbaum, 1988). Expert systems are computer programs intended to simulate the inference processes in expert judgment by the manipulation of rules with an "inference engine." Such rule-based systems proved to be extremely brittle in the sense that they produced nonsensical conclusions in situ-

ations not within the very narrow focus of their interest. Moreover, rule-based systems were ineffective in cases that were merely similar or analogous to those anticipated by the rule designers, and they generally did not incorporate any cognitively plausible strategic problem-solving knowledge. Some researchers have been attempting to address the brittleness problem by building systems having huge numbers of facts and rules; most notable among these is D. B. Lenat's Cyc project (Lenat & Guha, 1990).

- Procedural problem-solving tasks such as logic puzzles or games are unrepresentative of human intelligence.

Even today, many AI researchers continue to use tasks such as the Eight Queens problem or the Towers of Hanoi game as the tasks their programs attempt to perform. Early researchers chose to study simple, formalized tasks like these because they were easier to do, and they hoped that the models developed for these tasks would work well in the real world—all that would be necessary would be to start reintroducing the complexity, moving from toy domains to the real world. It has turned out that in many cases, trying to do this completely invalidated an algorithm that seemed quite reasonable at first glance.

Unfortunately, much research in cognition, and particularly in intelligence assessment, still relies on the assumption that logic puzzles are a good measure of general intelligence, or that models designed for such tasks are representative of the mechanisms of human intelligence.

- Mechanisms of intelligence are centrally dependent on the content of knowledge. Knowledge is not merely "grist" for the mind's mill.

Given the immense difficulty of reasoning in the simple domains we have been discussing, we must wonder how humans successfully reason in complicated, changing, poorly understood, or even adversarial environments. In attempting to write programs that can understand natural language, reason about people's actions, perceive in real time in noisy environments, and perform other authentic real-world tasks, the lesson has been driven home to researchers that what makes humans able to func-

tion in the real world is very large amounts of knowledge (Minsky, 1974).

This great need for knowledge first became clear in natural language understanding tasks. By the 1970s, programs had been written that could answer questions about simple texts, resolve anaphoric references, and so on. However, consider the following sort of utterance: “John went to a restaurant. He got a hamburger. He paid the check and left.” Did John eat? What did John eat? Why did he pay the check? It became clear that, in order to answer these kinds of questions, a machine or a human had to have a very detailed, ordered description of what one did in a restaurant, or on a bus, or when using the washing machine, or talking to the boss, or almost any situation where one had frequent, similar interactions. These structured descriptions were identified by Schank and Abelson (1977) and were called *scripts*.

Enormous mental resources would be required to plan our everyday actions and understand complex situations if we were to use the approach of mentally combining general rules and operators. The difficulty humans have coping with truly novel situations is good evidence that we ordinarily do not perform such chains of logical operations. Fortunately, we do not have to; almost every action that we are called upon to perform in the course of our daily lives is like one we have performed before. We do not need to figure out how to get food in a restaurant, dial a phone, or find our way around town. We usually are able to recall a plan and execute it almost unthinkingly. At most, we may have to adapt some familiar plan in some straightforward manner.

What makes people intelligent is not their ability to perform artificial problem-solving tasks in toy domains. What makes people intelligent is their ability to perform well in the human tasks that they must perform every day. Are they good at perceiving the motivations of others? Can they outwit their competitors? Can they learn from their mistakes and avoid similar ones in the future? Can they communicate well?

These tasks rely on the knowledge a person has, and on the organization and accessibility of that knowledge. To put it another way, it is the *stories* a person knows, and the ways in which those stories are

labeled and *interconnected* in memory, that is most important. The organization of this knowledge is both the cornerstone and the measure of human intelligence.

WHAT MAKES HUMANS INTELLIGENT?

What follows is a discussion of human intelligence from an AI perspective. This view of cognition is driven largely by one question: What have we had to build into our machines in order to make them do the things humans do?

It has become clear that only through the use of large amounts of real-world knowledge can we hope to have our machines behave in intelligent ways. This lesson has given rise to a view of human intelligence that is different from mainstream thinking in psychology. It is a view that acknowledges the central role of human memory and focuses on the way that knowledge is stored, organized, and retrieved from memory.

Here are some of the tasks that, from an AI perspective, form the essence of thinking.

Case Indexing and Retrieval. What is required in order to use knowledge? The approach of recalling earlier experiences to solve problems, as opposed to logical rule-based approaches, is called *case-based reasoning* (or *CBR*; see Riesbeck & Schank, 1989). In CBR, first a relevant *case* is recalled; a case can be a problem solution, a story, or any other experience stored in memory. Second, the retrieved case is adapted as necessary to the current situation. But these two steps are in themselves quite large problems.

How do we find a relevant case, with all the events of our lives and all of our personal stories to choose from? Oddly, much of the human memory research in cognitive psychology, which is a large and thriving field, has had very little to say about how we “get reminded” of things.

Our stories must somehow be labeled so that they can be retrieved. What is this labeling like? Consider a couple of examples.

In the first example, you remark on the interesting style of a teenager’s sandals, and, to your surprise, the teenager appears to be deeply insulted by your remark. Since this situation involved sandals, one label for this case could be *interesting sandals*. More likely, though,

the most significant feature of this situation was the teenager's reaction, which might cause you to use the label *caution about teenage hypersensitivity* to index this case.

In the second example, a salesperson goes on a sales call to a client named George before adequately preparing for the meeting. Two possible labels for this case might be *meeting with a person named George*, or *failing to make a sale*.

Why, in these examples, are the second labels preferable? The answer is that those labels are more potentially *useful*. Those labels will enable retrieval of those cases in new situations where one's goals are similar to the previous situations. Suppose the *teenage hypersensitivity* label is used to index the case; then, when in a situation where you are about to say something to a different teenager about his singing, you just might be reminded of that previous experience and act more wisely in this new situation. The *sandals* label, on the other hand, would not help you to retrieve the prior case in this new situation. Generally, labels are designed to help us to function more effectively in the future, and thus typically will be constructed in terms relating to our goals and the relevant circumstances of the situation.

On the other hand, in principle, anything can be a label. Sometimes we label cases based simply on interesting or unusual features, as, for example, when we see a colossal tree or hear a very witty expression. Another kind of label is created in anomalous situations that we did not anticipate or cannot explain. For example, we might create a label for the time when we drove to work and were stopped at all fourteen traffic lights on our route; most of our other trips to work are lost in a blur, but anomalous trips are typically more memorable.

Which labels are used to index a case is dependent on the level of expertise a person has in a given domain. A significant part of expert knowledge is knowing what kinds of features are likely to make good labels for cases. For example, when an engineer and the engineer's attorney friend watch a man trying to hoist a beam using a block and tackle, the attorney would tend to label the case in terms of "surface" features, such as *construction worker* or *ropes and pulleys*, while the engineer would label it in terms that are relevant to actually solving the worker's problem, such

as *problem using $F = ma$* , or *limited strength of nylon ropes*. Part of what makes engineers experts is how they have labeled the cases in their memory. They know, through experience, that the fact that the scenario contains a construction worker is likely to tell you very little about how to solve the problem. However, in another domain, such as labor law, the presence of a construction worker might prove of great importance, and we would expect the attorney who is an expert in labor law to label stories so that he or she could find stories about construction workers when necessary.

The more ways we have labeled a story, the more ways it might serve us in thinking about new situations, and the more intelligently we can handle those situations. Returning to the case of the unprepared salesman, *failing to make the sale* is not the only useful label. As the salesperson mulls over this experience, we can imagine him or her affixing new labels to the experience, such as *not understanding the client's industry*, *failing to memorize our company's sales pitch*, *knowing what competitors are saying*, and *arriving too early*. These are the central activities of intelligence: mulling over events, considering them from new perspectives, selecting the most important features, and, in short, creating good labels for the story. In CBR, this is referred to as the *indexing problem*.

In summary, the problems of retrieving appropriate cases from memory are not so very tough, if you have "done your homework." If the cases in memory are richly and appropriately indexed, it is no great trick to find a relevant story from memory. This would suggest that there is a habit of mind that might tend to make a person respond more intelligently to his or her environment. The habit of considering events from different perspectives and looking for interesting or important features will serve to make more appropriate and useful cases available at times when they are needed. Intelligent people generally have this habit of mind; interestingly enough, it could well be that it is the habit that makes them intelligent rather than the other way around.

Learning New Indices. All of this leaves open an important question: Where do all these labels come from?

Most labels are taught to us. When we are children, we witness an event, and an adult supplies us with labels to attach to that event. We see someone act

badly, and someone says to us, "Don't mind them, they're just jealous." We now have a label to apply to this story. We have, so to speak, created a "bucket" where we can keep stories about jealousy.

Occasionally, people will see a situation and create a new way of understanding it for themselves. A situation occurs that has no easy explanation, no ready label for us to use. Suppose you go to visit a friend who, you have learned, is dying. You may go there expecting to have an experience you will label *a time someone was miserable*, or even *a time I was miserable*. But suppose when you get there your friend is extremely happy about the whole affair, and has invited you over to celebrate. You have an opportunity to say goodbye and to tell your friend how much he has meant to you, and you make your peace with him and leave feeling pretty good. Nothing in your life has prepared you for this event. There is no story from memory that you can compare it with. Aside from vague, superficial indices like *story about death*, what meaningful, general indices will you use? You may adapt an old index, like *someone dying*, adding the features that make this situation unique: *time that someone was dying, but was glad of it*.

But this label is only descriptive of the circumstances. It will help you recall the story when someone else is dying and is glad of it. But it will not be much use to you in other situations that may be similar to this experience in interesting ways. You have come across an anomaly, someone who is not unhappy about dying. It would be much more helpful to you later on if, instead of simply categorizing the story to yourself, you manage to explain the anomaly; the explanation could then serve as a label for this case.

Explanation. Sometimes, the explanation for the event is not really an explanation at all. "He's obviously insane," you might say of your friend. In other words, there *is* no reasonable connection between his situation and his mental state. In this case, you have "explained away" the anomaly more than you have really explained it. However, even this explanation makes the story more useful to you. You might now use labels like *the effects of madness*, or *crazy people I have known*.

You might use your memory of this story to help you cope with the behavior of other people you perceive as irrational. But you won't gain much explana-

tory power until you start theorizing about the causality behind this anomalous behavior. There are two effects of this activity. First, you are discovering new ways to index the story, and thus are gaining a better understanding of it. Second, you are preparing yourself to understand new situations that are somehow similar or related to this behavior.

Of your dying friend, you might say, "He has determined to enjoy the time he has left to live," or "He has elevated death to a desirable status, to avoid feeling grief or fear." The correctness of your hypotheses is not the issue here; what is important is that you are creating new structures in your brain that might help you to understand situations that occur in the future.

When we observe an event that is unlike any we have seen before, it is actually quite rare for us to create an explanation by reasoning from scratch. Most of the time, we simply adapt an explanation we have heard before. Both of the explanations given above are common stories in our culture. Even if you had never applied them to death before, you had almost certainly formed explanation patterns like *enjoying what little you have*, or *being in denial*. Or course, all indices have some beginning. Even if we gain almost all of them as children by listening to the explanations of the adults around us, someone had to invent each and every one of them. Though not all of them are necessarily true or consistent, there are many thousands of commonly held theories and ideas that are culturally transmitted: *diseases are caused by microorganisms*; *men make better leaders*; *centralizing the provision of public goods is advantageous*; *natural selection causes the fittest to flourish*; *screws are removed by twisting*; *young people should marry within the faith*; and so on. None of these labels has existed forever; each was invented and subsequently disseminated.

Like mulling over events and thinking about them from multiple viewpoints, noticing anomalies and finding reasonable explanations for them is an essential part of what intelligent people do. But it appears likely that explanation makes us intelligent, just as intelligence gives us the ability to create explanations.

Here, then, is another habit of mind that can influence intelligence. By looking for anomalies in the events we experience, and by creating explanations for them (as opposed to simply explaining them away), we make ourselves more capable of thinking about new

events. With a larger arsenal of explanation patterns, with the events in our memories indexed in ways that are deeper, more general, and more incisive, we will be better equipped to think more effectively. We thereby make ourselves more intelligent.

Planning. One thing all intelligent entities must do is to plan. The simple act of deciding what to do next to achieve our goals—both the trivial ones and the important ones—is another hallmark of intelligence. Like many aspects of intelligence, it seems at first glance to be astonishingly difficult, and it is true that it is something that only intelligent people seem to do very well.

Although planning has been studied by AI researchers for a long time, only recently are there any models of planning which are suggestive of how humans plan. As mentioned previously, traditional models of planning feature very generic, knowledge-poor mechanisms. A human trying to use these methods would need to keep hundreds of conditions and plan steps in his head at the same time. How do humans actually plan? How do we create sequences of actions which satisfy our goals, when it is so difficult? The answer is, we cheat.

Here, as before, the answer lies in case-based reasoning. When we need to solve a problem—for example, the problem of obtaining a diamond necklace—we do not examine and sort through our theories and rules about gemstones, stores, physical location, and proximity. Instead, we retrieve and apply the standard “buying” plan: we obtain money, go to the jewelry store, show the clerk the money, ask him to hand over the necklace, and give him the money. Or if we don’t try to achieve our goal with the buying plan, then instead we’ll instantiate some other standard plan, such as stealing, handcrafting, borrowing, or convincing an acquaintance to give us one. We do not need to figure out a sequence of actions. We can just find an appropriate plan from memory, adapt it as necessary to the new situation, and execute it.

From the time we were told by our parents to ask for what we wanted instead of simply crying because we didn’t currently possess it, to the time that we learned ways to buy things of different value, we have been learning plans for obtaining objects we want. It is not surprising, then, that we get pretty good at it. The difficulties that arise in this kind of exercise are

no longer the type associated with traditional planning models. We do not get stuck trying to choose from among the many possible simple actions we could do next. But we might get stuck knowing which plan is the most appropriate one to execute at a given time, and when to abandon one plan (such as when we realize we don’t have nearly enough money to buy the necklace) in favor of another one (such as borrowing). The indecision that can plague us at times like these is good evidence that we do not construct a new, detailed sequence of plan steps for our next action when we already have perfectly good plans in memory which we can follow.

It is only when we must radically adapt a plan to make it fit a new situation, or when we must decide between plans, or reason about them, that our basically poor ability in planning becomes clear. This is not to say that people are not intelligent; it is just that the task of planning from scratch is extremely difficult. This is not to say that we never create new plans. However, it is a relatively rare event. Mostly, we adapt an old plan, or execute it as it is. Everyone has seen examples of people executing plans that were not really appropriate, solely because they had always done it that way.

Oddly, this view of planning as the execution of precreated plans from a library of plans did not originate with planning researchers. It was first postulated for programs that interpreted natural language, such as the PAM program (Wilensky, 1978). What was a program to make of an utterance like, “Willa was hungry. She grabbed a restaurant guide. She got in her car”. How could such a program be able to answer questions like, “Why did Willa get in her car?”

A human knows that there are many possible plans for satisfying one’s hunger, and that one of them is to go to a restaurant. We know that one standard plan for being at a place requires the agent to get in the car and drive it to that place. Without this kind of knowledge, there is simply no way a computer program, or a person, could understand much of what is said and done around it.

Communicating. So far, we have claimed that thinking relies on vast amounts of knowledge, most of which we learned from others. Wouldn’t this require that we spend a great deal of time and energy getting knowledge from people and giving it to others? In fact,

we do; it is the most uniquely human cognitive activity. We spend a large percentage of our waking hours talking, listening to other people's stories, and telling our own. Go into a public place, and watch the people come and go. Except for those who are alone, the rest talk, talk, and talk. What on earth are they saying? Why do they do this?

One reason is clearly to transmit information from one person to another. Sometimes we can see when someone is in need of information, and we can give it to them. We might say, "That door is a little sticky, you have to pull it real hard," or, "The customer thinks his order has already been shipped!" When we are in need of information, we might simply ask for it, and someone who knows will typically be willing to tell it to us.

But there must be more reason for communicating than that, or the world would be a considerably quieter place. Our heads are simply full of stories, and we love to tell them. We like to tell people when we've had a day that was good, bad, or even just different. We like to tell them where we're from, how we got here, and what we think of this or that event, situation, or person. We tell what happened to us, what's happening to us, what we hope will happen to us, and what we fear will happen to us. And then we start on our friends, and tell what's happening to them.

We tell stories to let people know who we are and what we are like, to let them know who we think *they* are, and what we think *they're* like. If someone tells us a story of something that happened to them, we are very likely to tell them a story about when something similar happened to us.

The human mind is stuffed with stories. In an important sense, stories provide us with our very identities. They are not only what we think about, they are what we think *with*, since we use them to recognize and interpret similar situations in the present. The most central activities of human intelligence involve integrating new stories into memory, connecting them to what we already know, labeling them for retrieval later, and relating them to our goals and our current situation. We seem to be constructed in such a way that we do this best when we are articulating our thoughts and experiences. This is another part of why we compulsively tell people our stories. When we are retelling a story, we notice things about it we had not

seen before, and we discard parts of it that do not seem important.

Telling stories is as natural as the way, for example, that chimpanzees groom each other. In fact, storytelling serves a similarly useful function. By telling our stories, we are refining them, elaborating them, and ordering them. By listening to our stories, people around us are not simply learning about the world for their own benefit. They are helping us to order our thoughts, index our cases, and understand our lives. By listening to and helping us to think about our stories, they are helping us to perform a basic maintenance task, making our lives significantly easier, and keeping us psychologically well adjusted. And, reciprocally, our listening to others' stories has a mutually beneficial effect for us.

Like mulling and explaining, storytelling is another habit of mind that prepares us to think better. By telling our stories, we crystallize our thoughts and formulate opinions. As E. M. Forster said, "How can I tell what I think until I see what I say?"

The collection of stories in a person's memory probably provides the best single indicator of the level of intelligence of the person. Intelligence is a function of both the number of different stories a person knows and the quality of the indexing of those stories. Quality of indexing, in turn, is a function of the labels used to index each story and of the functional effectiveness of those labels—in other words, how regularly the right story is retrieved in the right circumstances.

THE FUTURE OF AI

Indexing and retrieving cases, learning new indices, explaining, planning, and learning and telling stories are all active research areas in AI, and there are many more besides. There is excellent work being done in perception, emotions, and common-sense reasoning, just to name a few areas. With all this work happening, it is natural to want to ask whether we will ever put it all together and create truly intelligent machines. What will AI machines of the future be like?

To answer that question, we must realize that AI programs are created artifacts. They are made to serve a purpose. To think about what AI programs will be like, we must decide what they will be for and what niches they are created to fill.

The results of much AI research will find their way into devices that we would not be disposed to call intelligent. A car that chooses a route home from work, follows the roads, avoids accidents, parks itself, and then wakes the passenger would be quite clever, and would require many techniques derived from AI. But we would not tend to grant it a point of view, beliefs, and so on.

But what about a program that performed tasks that are less mechanical? For example, governments and large corporations generate vast volumes of information, and classifying and organizing it is a tremendous task. A machine that could categorize and cross-reference documents would be of tremendous assistance. Now suppose that the machine did more than that. Suppose that it generated warnings when it noticed dangerous confluences of events, signaled when it detected an opportunity to use a previously successful stratagem in a new situation, or provided an explanation or analogous case for a puzzling new set of circumstances.

The first truly intelligent machine is unlikely to be anything like the human-sized, chattering robots depicted in movies, because having such robots is not a pressing societal need. Rather, it could well be that the first really intelligent machine will live in large government or corporate computer systems. The program could serve as a smart, experienced, continually learning, and untiring "analyst" or "consultant" that silently reads reports, news wire services, statistics, and so on. Such a machine would need to have a memory for events, much in the way a human has, and it would need to organize its memory and usefully label what it observes. Since its purpose would be to extract information from what it is reading, it would need to formulate opinions and to express them. It would need to determine when a person needed to hear a particular story, and to tell it, and in some cases to educate the person. It would need to be able to find answers to questions that are put to it, whether these queries are for factual information, opinions, or explanations of its own reasoning. Further, it would be necessary for the machine to be able to frame questions of its own and understand the answers. In short, it would be able to converse, knowledgeable and extensively, within its areas of expertise.

Except on abstruse philosophical grounds, it is difficult to say that such a program would not be intelligent. And yet the properties described do not seem impossible, in principle, to program into a machine.

Nonetheless, some people would dismiss the idea as fantasy. But the possibility of such a machine is not the paramount issue. In building it, or even in just attempting to build it, we will reap great technological benefits. More important, we will come closer to an understanding of the workings of the human mind.

(See also: CONNECTIONISM.)

BIBLIOGRAPHY

- FEIGENBAUM, E. A. (1988). *The rise of the expert company*. New York: Times Books. Describes expert systems and the varieties of their application.
- LENAT, D. B., & GUHA, R. V. (1990). *Building large knowledge-based systems: Representation and inference in the Cyc Project*. Reading, MA: Addison-Wesley.
- MINSKY, M. (1981). A framework for representing knowledge. In J. Haugeland (Ed.), *Mind design*. Cambridge, MA: MIT Press.
- NEWELL, A. (1990). *Unified theories of cognition*. Cambridge, MA: Harvard University Press. Describes the SOAR architecture.
- NEWELL, A., & SIMON, H. A. (1963). GPS: A program that simulates human thought. In E. A. Feigenbaum & J. Feldman (Eds.), *Computers and thought* (pp. 279–293). R. Oldenbourg KG.
- RIESBECK, C., & SCHANK, R. C. (1989). *Inside case-based reasoning*. Hillsdale, NJ: Erlbaum. A broad sampling of implemented CBR systems.
- SCHANK, R. C., & ABELSON, R. P. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, NJ: Erlbaum.
- WILENSKY, R. (1978). *Understanding goal-based stories*. Ph.D. dissertation, Yale University.

ROGER SCHANK

ARTISTIC ABILITY: CHILDREN'S DRAWINGS Children's drawings have been of interest to psychologists and educators for over 100 years. During this period, investigators have made concerted efforts

to delineate a series of developmental stages that would correspond to children's level of conceptual development and thus link the perceived deficiencies in the childish drawings to their mental maturity. Despite significant differences in their positions, investigators shared the common assumption that graphic development progresses toward an ideal norm of realism in art, and that deviations from photographic likeness was a mark of conceptual deficiency.

More recently, students of child art have challenged this position and provided empirical evidence for a very different view of the developmental progression in drawing.

THEORETICAL PERSPECTIVES

R. Arnheim, the most influential theorist in this field, has formulated a theory of representation that applies equally well to children and adults. In his view, representation from its very beginnings rests on the invention of forms that are structurally or dynamically equivalent to the object. Representation does not aim for one-to-one correspondences and artists do not aspire to copy the original, an impossible task given the intrinsic differences between the properties of a two-dimensional and a three-dimensional medium. By its very nature, the drawing medium with its specific tools tends to encourage shapes made of lines and contours. These shapes comprise the basic constituents out of which a complex graphic language will evolve, beginning with simple shapes and relations and gradually developing into more complex and differentiated systems of representation.

This view of children's drawings as representations opened new avenues for investigating the relationship between children's knowledge of an object and its rendition. Studies on the effects of tasks, media, and instructions revealed a much richer picture than previous theories had predicted, and the notion that a drawing might be seen as a conceptual printout of the child's mind was laid to rest (Freeman, 1980; Golomb, 1973, 1974, 1992). Instead, a new appreciation of the child artist emerged and with it came the understanding that the child who draws a circle, equips it with facial features, and calls it a man, does not mistake it for a human of flesh and blood, but merely views the

figure as an acceptable equivalent. The child evolves a rule system that holds that a unit on paper, for example, the circle with facial features, is equivalent to another unit, a human or an animal. This conception of equivalence does not reflect an inadequate percept or a faulty concept of the object. On the contrary, it is a mark of the symbolizing propensity of the human mind to establish correspondences on the basis of general qualities.

Thus, the invention of a meaningful graphic language rests on an understanding that symbols represent, and that they are not to be mistaken for the actual object they refer to, that is, that symbol and referent are to be distinguished. From this perspective, the child's first drawings are original productions in the sense that they are not derived from an existing model (see Figure 1). Humans are not made of circles, dots, and lines, and one might best view the child's rendition as its own invention.

THE COURSE OF DEVELOPMENT

The beginnings of artistic development can best be seen as a transition from perceptual-motor action to representation. Depending on the availability of paper and crayons, a slate, rock, or sand, toddlers and preschoolers tend to explore the effects their gestures can have on a flat surface, especially when they leave clearly visible marks on paper. These marks can vary from rotational whirls that mostly record motor action, with limited interest in the ensuing shapes and edges of the paper, to zigzags and more carefully placed line formations. A major milestone in artistic development is reached when the child discovers that forms look like something, and can represent a familiar object. Unlike the early scribble patterns, representation is a symbolic activity where marks, be they dots, lines, or contours, point beyond themselves to another realm of meaning.

With the advent of graphic representation during the child's fourth or fifth year, a new level of awareness emerges that leads to often tireless efforts to represent objects or events that are of significance to the child. Given the three-dimensional character of most objects, the child's drawn version does not stand in a simple relation to its model. A child's first drawings of

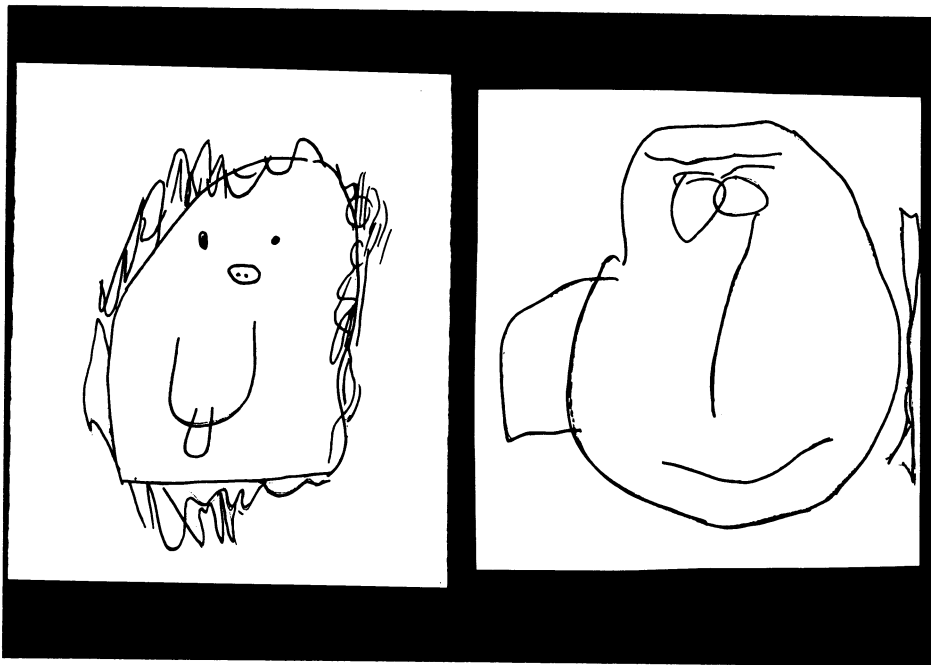


Figure 1

Global human figures. The circle with facial features represents one of the earliest graphic models of the human figure. (Girl, 3-3; boy, 3-8).

humans and animals consist of global units, circles or oblongs, that encompass the facial features. Depending on experience with the medium and the motivation to draw, the simple figures will soon undergo differentiation and come to reflect the child's interest in more diverse forms, detail, size, proportion, orientation, action, dimensionality, views, color, and the relationship among all the elements of the composition.

The developmental course, especially in its early phases, appears to be quite universal, and graphic problem solving follows predictable patterns. It is important to note, however, that the development of increasing complexity and skill in pictorial representation does not have a specific timetable nor a unique endpoint, for example, realism in art. A stage conception of development joined to a rigid timetable cannot account for the rapid progress many children make in the graphic medium when their motivation to explore can lead to a remarkable differentiation within a single session. It cannot account for a child's diverse representational models when working on different themes and with different media. A narrowly conceived stage conception of development also fails to account for

children who are precociously gifted in the domain of drawing. The case of Eytan (Golomb, 1992) presents a normally developing child who progressed at such a rapid pace that, at the age of three and four years he had mastered, on his own, multiple three-dimensional strategies for the depiction of vehicles and scenes of interest to him (see Figure 2). The case of Nadia, an autistic child (Selfe, 1977), is most remarkable in that an otherwise retarded young child showed unusual talent in her portrayal of animals in motion, executed with great fidelity and at a very advanced level.

Thus development of child art can be seen as a problem-solving process that requires solutions to an intractable problem, namely, how to represent a three-dimensional object on a flat surface. In general, this problem-solving process begins with general-purpose forms and simple relations and, depending mostly on the child's experience and active experimentation with the medium, becomes increasingly complex and differentiated in terms of the elements that are depicted, the forms that are selected, the relations that obtain among them, and the views that are considered. The findings from recent studies indicate that a drawing is



Figure 2
 This drawing of a truck depicts its multiple sides (front, side, top) by using parallel oblique lines. (Eytan, age 3-7).

not to be seen as a printout of the artist's concept of the object, an outdated assumption that is the basis for the Goodenough-Harris Draw-a-Person Test (1963) as a test of verbal or conceptual intelligence. Instead, drawing development is best understood as the artist's search for forms of equivalence that can stand for the object. In the beginning phases of this developmental course, the rules that determine the early graphic models are of a universal order. At a later time, especially during the elementary school years and beyond, the influence of culturally dominant models, for example, of cartoon characters, will become more prominent, and some children will become interested in matching or copying the graphic models available to them.

BIBLIOGRAPHY

- ARNHEIM, R. (1974). *Art and visual perception*. Berkeley: University of California Press.
- COX, M. (1992). *Children's drawings*. London: Penguin.
- FREEMAN, N. L. (1980). *Strategies of representation in young children*. London: Academic Press.
- GARDNER, H. (1980). *Artful scribbles: The significance of children's drawings*. New York: Basic Books.
- GARDNER, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- GOLOMB, C. (1973). Children's representation of the human figure: The effects of models, media, and instruction. *Genetic Psychology Monographs*, 87, 197-151.
- GOLOMB, C. (1974). *Young children's sculpture and drawing*. Cambridge, MA: Harvard University Press.
- GOLOMB, C. (1992). *The child's creation of a pictorial world*. Berkeley & Los Angeles: University of California Press.
- GOODENOUGH, F. L. (1926). *Measurement of intelligence by drawings*. New York: Harcourt, Brace & World.
- HARRIS, D. B. (1963). *Children's drawings as measures of intellectual maturity*. New York: Harcourt, Brace & World.
- KERSCHENSTEINER, G. (1905). *Die Entwicklung der zeichnerischen Begabung*. Munich: Gerber.
- LIQUET, G. H. (1927). *Le dessin enfantin*. Paris: Alcan.
- PIAGET, J. (1928). *Judgment and reasoning in the child*. London: Routledge.
- PIAGET, J. (1951). *Play, dreams, and imitation*. London: Routledge.

- PIAGET, J., & INHELDER, B. (1956). *The child's conception of space*. London: Routledge.
- SELFE, L. (1977). *Nadia: A case of extraordinary drawing ability in an autistic child*. London: Academic Press.
- WINNER, E. (1982). *Invented worlds: The psychology of the arts*. Cambridge, MA: Harvard University Press.

CLAIRE GOLOMB

ASIAN AMERICANS In examining issues related to Asian-American intelligence, some definitions and background information regarding who falls within this racial/ethnic classification make a helpful beginning. The category of Asian Americans includes many diverse subgroups such as Chinese, Japanese, Filipinos, Koreans, Asian Indians, Vietnamese, Laotians, and Cambodians. Each subgroup has a unique cultural background, language, religion, and history in the United States. Many of the studies on the intelligence of Asian Americans (Lynn, 1987) have been conducted primarily with samples of Chinese and Japanese people.

Asian Americans have been identified as the "model minority." They obtain high scores on standardized test instruments (i.e., measures of mathematical abilities), obtain higher grade-point averages than their peers in certain regions of the United States, spend more time on homework, take more advanced courses in high school, graduate with more credits than other American students, and have higher completion rates for high school and college than do Caucasian students (Brand, 1987).

Asian-American populations tend to be concentrated in California, Hawaii, and New York. Populations in these states tend to have higher incomes relative to the rest of the United States because the residents experience a higher cost of living than the national average. In addition, Takaki (1989) reports that Asian-American families tend to have more persons working per family than Caucasian families. Therefore, higher incomes are reflective of more workers within each family. Further examination of these results reveals that Asian-American men (in this study, Japanese) earned comparable incomes to their Caucasian counterparts, and they also had more years of education and worked more hours. Income discrep-

ancies between other Asian-American groups and Caucasians were also noted.

D. Brand (1987) reports that many Asian Americans do not agree with the "model minority" myth, as it leads to a false stereotype of a group that is experiencing no problems. Although there are examples of individuals within the Asian-American community who have excelled and achieved far beyond their peers, a uniformly high level of success does not typify this diverse group of people.

Although Asian Americans are not faring as well as reports would have us believe, many researchers have attempted to explain the factors that account for the supposedly uniformly high achievement of this racial/ethnic group. S. Sue and S. Okazaki (1990) cite two hypotheses: (1) hereditary/genetic differences in intelligence, and (2) Asian cultural values that emphasize educational pursuits. The following discussion highlights issues related to these two hypotheses. The areas to be covered include Asian-American IQs and profiles of abilities, explanations for visual-spatial strengths and relatively lower verbal abilities, and explanations for high achievement.

ASIAN-AMERICAN IQ AND PROFILES OF ABILITIES

Prior to addressing the IQ of Asian Americans, it is important to acknowledge the difficulties in obtaining and interpreting data regarding the intelligence of these racial/ethnic subgroups. J. Flynn (1991) has examined Asian Americans' IQ studies conducted over a period of approximately fifty years. He cautions readers regarding the interpretation of a global IQ for the Asian-American community. In particular, he cites the diversity of the Asian racial/ethnic subgroups. J. Hsia (1983) suggests that one cannot generalize information about the cognitive abilities of one Asian subgroup to another without consideration of experiences and culture.

Others have indicated concerns as research addressing the IQ of Asian Americans has suffered from the "absence of adequate norms, nonstandardized test translations, inadequately defined criterion behavior (i.e., that which is being measured), etc." (Chin, 1983, p. 102). Chin suggests that evaluators use caution in

interpreting IQs of Asian-American children. Although this concern is especially true for immigrant children, more subtle cultural influences may be noted for second- or even third-generation Asian-American children.

Despite these differences between racial/ethnic subgroups and methodological problems, numerous studies have explored the intelligence of Asian Americans. Historically, some studies have indicated that Asian Americans obtain higher overall IQs in comparison to other racial/ethnic minority groups. Flynn has criticized these studies (1991) and has indicated that the IQs reported were inflated because of usage of inappropriate norms. Adjusted Asian-American IQs are commensurate with overall IQ scores obtained by Caucasian Americans.

In his review of sixteen studies conducted between 1939 and 1985, including 11,373 Japanese and Chinese Americans aged 8 to 25 years, Flynn (1991) notes that the average overall IQ ranged from approximately 101 to 122, but scores from 91 to 99 were adjusted to reflect appropriate norms. Adjusted nonverbal IQ was estimated between 94 and 102, and adjusted verbal IQ ranged from 86 to 98. The Asian-American IQ estimates in one study dropped 18 to 20 points after adjustment. These adjustments led Flynn to conclude that Asian-American IQs may be inflated in various studies.

Findings from The Urban Institute indicate that the patterns of Asian-American IQ scores is "a relatively simple one of lower-than-average IQs in the early years (of the 20th century) and higher than average scores in the later years" (Sowell, 1978, pp. 212–213). Sowell also reported that Asian-American children scored higher than white children on particular subtests that did not involve a language component, suggesting that language plays an important role in the measurement of intelligence using traditional tests. Flynn (1991) later adjusted the IQs obtained in this study estimating that the overall Asian-American IQ is approximately 98.

In addition to examining the overall IQ of Asian Americans, numerous studies have explored their ability patterns in comparison with other racial/ethnic groups (see ETHNICITY, RACE, AND MEASURED INTELLIGENCE). With respect to the ability patterns demonstrated by Asian Americans, studies indicate that they

have higher visual–spatial abilities in comparison to verbal abilities.

For example, Wing (1980) compared the abilities of Asians, whites, blacks, and Mexican Americans on a series of tests measuring the following abilities: verbal, judgment (e.g., problem solving), induction (e.g., analogies), deduction (e.g., reasoning from general ideas to specific situations), and number (i.e., mathematical abilities). Findings indicated that whites and Asians have higher averages than the other groups. Profiles of whites, blacks, and Mexican Americans tended to be flat. Blacks performed slightly better on verbal tests relative to the other areas assessed (judgment, induction, deduction, and number). Asian Americans did better on induction, deduction, and number (measures assessing more nonverbal abilities) in comparison with verbal and judgment measures (measures assessing verbal abilities).

In a study comparing Americans of Japanese Ancestry (AJA) and Americans of European Ancestry (AEA), A. Marsella and C. Golden (1980) found that AJAs tend to score higher on tests measuring nonverbal abilities in comparison with verbal abilities. AEAs, on the other hand, were found to score higher on tests requiring verbal abilities. Marsella and Golden conclude that although the structure (general make-up) of cognitive abilities of AJAs and AEAs are similar, they may differ with respect to the specific skills used to resolve problems. For example, when facing various tasks, each group may solve the problems using different subsets of cognitive abilities.

P. VERNON (1982) notes in his review of the literature conducted on the abilities of Asians in North America the similar findings for Chinese-American children. As early as the 1920s, they demonstrated lower scores on verbal intelligence tests and verbal achievement in comparison with their white peers. They also scored equally or higher than whites on nonverbal intelligence measures. Early studies indicated that the verbal–nonverbal discrepancy in scores was "quite considerable." More recent studies have indicated that average IQs for Chinese-American children are approximately "97 for verbal tests and 110 for nonverbal and spatial tests" (Vernon, 1982, p. 28). Note that no adjustments were made in Vernon's review of the research on the IQ of the Asian subgroup studied.

The pattern of high visual-spatial abilities and relatively lower verbal abilities is consistent throughout much of the literature addressing Asian-American intelligence (Chin, 1983; Hsia, 1983; Nagoshi & Johnson, 1987). Vernon (1987) writes:

It is interesting that all of the groups of Mongoloid (Asian) origin that I have studied, including Chinese in Taiwan, Hong Kong, Hawaii, U.S., and Canada, Japanese at home or abroad, Eskimo and Native Indian, show this tendency to score more highly on nonverbal and spatial than on verbal tests. The objection that all groups tested in English were handicapped by lack of familiarity with the English language, is contradicted by the fact that many Japanese and Chinese in America or Canada have been living in English-speaking environments for several generations, and mostly speak English at home, at school, or in their jobs. But although they have caught up to about the white average on verbal tests, they are still about 10 points better on spatial tests [p. 385].

Studies have also indicated that Asian Americans tend to have high mathematical abilities (Hsia, 1983; Vernon, 1982). Vernon notes that Chinese students are superior in mechanical arithmetic but not in problem or applied mathematics (p. 28). Among a group of Chinese-American students, males demonstrated a large discrepancy between their quantitative and verbal scores. They also showed specific strengths in quantitative abilities. On the other hand, the females in the study demonstrated a relatively closer balance, verbal abilities being "slightly favored" (Sue & Kirk, 1973, p. 472). These findings suggest the possibility of gender differences in Asian-American abilities.

EXPLANATIONS FOR THE VISUAL-SPATIAL STRENGTHS AND RELATIVELY LOWER VERBAL ABILITIES

Various explanations have been suggested to explain the visual-spatial strengths in comparison to the relatively lower verbal abilities of Asian Americans. These include evolution and neurological differences, and social and cultural explanations.

Evolution and Neurological Differences. R. Lynn (1987) hypothesized that the pattern of abilities consistently noted for Asians and Asian Americans cannot be explained by environmental factors alone.

He proposed an evolutionary and neurological explanation. He suggests that during the Ice Age, the extreme cold created a selection pressure for increases in *g* (general intelligence) and visuospatial abilities. Asians became dependent upon hunting for their food and visual-spatial abilities were vital in hunting. The "enhancement of visuospatial abilities in the Mongoloids (Asians) took place at the expense of verbal abilities . . . verbal abilities were sacrificed to permit an increase in visuospatial abilities" (Lynn, 1987, p. 833).

This pattern of abilities led to the development of distinctive features in the neurology of the brain. Lynn cites literature indicating that verbal abilities are located in the left hemisphere and visual-spatial abilities located in the right hemisphere. Thus, the relatively higher scores obtained on visual-spatial and nonverbal reasoning tasks suggest that Asian Americans demonstrate strengths in right-hemisphere abilities. Lynn notes that the Asian pattern of abilities and the neurology of the brain must have a "genetic basis which may possibly be enhanced by environmental processes" (1987, p. 837).

Cultural Explanations. Other researchers report that many Asian groups value nonverbal methods of communication. They rely on what is implied in the interaction rather than on what is expressed. For example, in the Japanese culture it is

adaptive for one to be more indirect and ambiguous. The Japanese also value implicit, nonverbal, intuitive communication, over explicit, verbal and rational exchange of information. Family and in-group members rely more on nonverbal cues and physical contact for real communications. (Hsu, et al., 1986, p. 320.)

Morsbach (1980) cites historical factors that may contribute to the Asian (in this case, Japanese) emphasis on nonverbal abilities. For example, in Zen philosophy the means to enlightenment is through meditation rather than verbal action or behavioral performance.

EXPLANATIONS FOR HIGHER ACHIEVEMENT

Flynn (1991) reports that Asian Americans (in this study, Japanese and Chinese) "achieve far beyond what their mean IQ would lead us to expect. . . . For ex-

ample, the post-war generation of Chinese Americans, those born from 1945 to 1949, had a mean IQ of 98.5 with Whites set at 100. But their achievements in terms of education, occupation, and income suggest an estimated IQ about 21 points higher than their actual IQ" (Flynn, 1991, p. 1).

Flynn further hypothesizes that Chinese Americans capitalize more effectively on their available pool of talent. He reports a similar phenomenon for Filipinos.

Numerous explanations have been suggested regarding the reasons for the achievements of Asian Americans. These reasons will be discussed in terms of culture, functional relativism, and biopsychosocial factors.

Culture. S. Sue and S. Okazaki (1990) summarize the cultural explanations for the achievements of Asian Americans as follows:

The most popular cultural view is that Asian family values and socialization experiences emphasize the need to succeed educationally. Largely on the basis of anecdotal and observational evidence rather than on empirical findings, investigators have identified the following values or practices in Asian families that promote educational achievements: demands and expectations for achievement and upward mobility, induction of guilt about parental sacrifices and the need to fulfill obligations, respect for education, social comparisons with other Asian-American families in terms of educational success, and obedience to elders as teachers [pp. 915–916].

Another explanation focuses upon the notion that Confucian ideals emphasize education. Thus, the philosophical/religious history of Asian-American groups promote the cultural importance placed upon educational mobility.

Sue and Okazaki also note that "culture is a concept that has been used to explain all phenomena, but one that is difficult to define and to test" (1990, p. 916). In their review of the literature investigating the cultural hypothesis, Sue and Okazaki conclude that values regarding hard work and the importance of study are meaningful predictors of achievement.

Examination of the achievements of Asian Americans is complex because of the multitude of factors that can have an impact on the analysis. For example,

Sue and Kirk cite research supporting the important role of the acculturation process. As "Chinese progressively become removed from their ethnic subculture and exposed to western influences, the norms and values of American society begin to guide and modify the orientations of the Chinese" (1972, p. 478). Brand (1987) reviews research indicating that as Asian Americans acculturate into the American mainstream, some demonstrate less interest in school and lower achievement. Hsu and colleagues (1987) note that many immigrant families tend to become acculturated after three or four generations.

Relative Functionalism. The academic achievements of Asian Americans do not derive only from cultural factors. Sue and Okazaki suggest that an ethnic group's status within a particular society has an influence on their achievement. Relative functionalism refers to the notion that Asian Americans have perceived or experienced barriers in upward mobility in jobs and careers not related to education. Jobs or careers that have restricted entry include entertainment, sports, and politics. According to relative functionalism, educational achievement has become the vehicle toward better jobs and career advancement. Thus, "when opportunities for upward mobility are limited or perceived to be limited in other areas, educational achievements should increase" (Sue & Okazaki, 1990, p. 917). Sue and Okazaki suggest that this finding is especially true for groups whose cultural values promote educational achievement.

Biopsychosocial Factors. In response to Sue and Okazaki's article regarding functional relativism, Fox (1991) suggests a linkage between the intelligence demonstrated by Asian-American groups and cerebral functioning. He notes a shared neuropsychological phenomenon between Asian Americans and Jews as both Asian languages and Hebrew are read and written from right to left. He observes that the right-to-left directionality has impact on the hemispheric stimulation of the brain. He hypothesizes that this directionality may "lead to increased and flexible cerebral functioning that affects intellectual performance" (1991, p. 878). Although Asian Americans may no longer speak or read their cultural language, "hemispheric proclivities may nonetheless be present" (1991, p. 878).

CONCLUSION

An answer to the question about Asian Americans' achievements beyond their IQ remains inconclusive. Intellectual indicators are not always available, however, as the primary groups studied are either Japanese or Chinese or both. As Flynn (1991) writes, "Unless scholars break the current taboos, and collect IQ data for Chinese, and Japanese, and Filipinos, and Vietnamese, and so forth, the usefulness of studies of Asian American IQ is probably at an end" (p. 59).

(See also: ETHNICITY, RACE, AND MEASURED INTELLIGENCE.)

BIBLIOGRAPHY

- BRAND, D. (1987, August). The new whiz kids: Why Asian Americans are doing so well, and what it costs them. *Time*, 130, 42-51.
- CHIN, J. L. (1983). Diagnostic considerations in working with Asian-Americans. *American Journal of Orthopsychiatry*, 53(1), 100-109.
- FLYNN, J. R. (1991). *Asian-Americans: Achievement beyond IQ*. Hillsdale, NJ: Erlbaum.
- FOX, D. (1991). Neuropsychology, achievement, and Asian-American culture: Is relative functionalism oriented times three? *American Psychologist*, 46(8), 877-878.
- HSIA, J. (1983). Cognitive assessment of Asian-Americans. In M. Chu-Chang & V. Rodriguez (Eds.), *Asian and Pacific perspectives in bilingual education: Comparative research* (pp. 123-152). New York: Columbia Teachers College Press.
- HSU, J., TSENG, W., ASHTON, G., MCDERMOTT, J. F., & CHAR, W. (1986). Family interaction patterns among Japanese-American and Caucasian families in Hawaii. In S. Chess & A. Thomas (Eds.), *Annual progress in child psychiatry and child development*. New York: Brunner/Mazel.
- LYNN, R. (1987). The intelligence of the Mongoloids: A psychometric, evolutionary, and neurological theory. *Personality and Individual Differences*, 8, 813-844.
- LYNN, R. (1991). Educational achievements of Asian Americans. *American Psychologist*, 46, 875-876.
- MARSELLA, A. J., & GOLDEN, C. J. (1980). The structure of cognitive abilities in Americans of Japanese and of European ancestry in Hawaii. *The Journal of Social Psychology*, 112, 19-30.
- MORSBACH, H. (1980). Major psychological factors influencing Japanese interpersonal relations. In N. Warren (Ed.), *Studies in cross-cultural psychology* (Vol. 2, pp. 317-342). New York: Academic Press.
- NAGOSHI, C. T., & JOHNSON, R. C. (1987). Cognitive abilities profiles of Caucasian vs. Japanese subjects in the Hawaii Family Study of Cognition. *Personality and Individual Differences*, 8, 581-583.
- SOWELL, T. (1978). Race and IQ reconsidered. In T. Sowell (Ed.), *Essays and data on American ethnic groups* (pp. 203-229). Washington, DC: The Urban Institute.
- SUE, D. A., & KIRK, B. A. (1973). Differential characteristics of Japanese-American and Chinese-American College Students. *Journal of Counseling Psychology*, 20, 142-148.
- SUE, D. A., & KIRK, B. A. (1972). Psychological characteristics of Chinese-American students. *Journal of Counseling Psychology*, 6, 471-478.
- SUE, S., & OKAZAKI, S. (1990). Asian-American educational achievements: A phenomenon in search of an explanation. *American Psychologist*, 45, 913-920.
- SUE, D., & SUE, S. (1987). Cultural factors in the clinical assessment of Asian Americans. *Journal of Consulting and Clinical Psychology*, 55, 479-487.
- TAKAKI, R. (1989). *Strangers from a different shore: A history of Asian-Americans*. Boston: Little, Brown.
- VERNON, P. E. (1982). *The abilities and achievements of Orientals in North America*. New York: Academic Press.
- VERNON, P. E. (1987). Cognitive and motivational differences between Asian and other societies. In S. H. Irvine & S. E. Newstead (Eds.), *Intelligence and cognition: Contemporary frames of reference* (pp. 377-392). Boston: Martinus Nijhoff.
- WING, H. (1980). Profiles of cognitive ability of different racial/ethnic and sex groups on a multiple abilities test battery. *Journal of Applied Psychology*, 65, 289-298.

LISA A. SUZUKI
TERRY B. GUTKIN

ATHLETIC ABILITY Research literature dealing directly with athletic ability and intelligence is very limited. However, some work has been done on athletic ability as it relates to academic achievement and on exercise and cognitive functioning.

ATHLETIC ABILITY AND ACADEMIC ACHIEVEMENT

The study of the relationship between sport (particularly athletics) and intelligence/academic achievement is fraught with methodological problems: (1) the exclusion of lower-achieving athletes who cannot meet eligibility requirements; (2) athletes who are viewed differently than nonathletes by instructors; and (3) the differential motivation to perform in athletes and nonathletes.

Given these restrictions, Hardman (1973) has addressed the relationship between athletic participation and intelligence. He reviewed sixteen studies, containing forty-two different adult athletic samples, that were tested with Cattell's Sixteen-Factor Personality Profile. The only consistent personality factor was factor B (mentally bright vs. dull). In all thirty-seven samples assessing factor B, the athletes' score was either average (5.0–6.0) or above average (>6.0). Hardman (1973) concluded that "whatever the reason, present results would indicate that the unintelligent do not choose organized sport as a form of recreation at the adult level . . ." (p. 97).

Fourteen studies on athletes' academic aspirations and achievement have been reviewed (Snyder & Spreitzer, 1983). There was a positive relationship between athletic participation and academic orientation. Compared to nonathletes, athletes had higher aspirations for going on to college as well as actual entry into college. The relationships were stronger for athletes of lower socioeconomic status who otherwise might not be disposed toward academic achievement. Furthermore, "star" basketball players had higher college aspirations, got more advice from coaches on whether to attend college, and perceived their coaches to be a greater influence than "starters" and "substitutes." The findings suggest that with the prestige and visibility associated with sport, athletes are given encouragement by parents, teachers, and coaches, and are members of the leading crowd which influences educational plans and expectations beyond high school (Snyder & Spreitzer, 1983, p. 132).

When it comes to actual academic attainment, the picture becomes more complex. Although some studies have found that athletes have higher grade-point averages than nonathletes, other studies have not sup-

ported this relationship. Spady (1970) showed that athletes dropped out of college more than athletes or nonathletes who were involved in other extracurricular activities (e.g., student government, interest clubs, drama). Furthermore, Landers and colleagues (1978) showed that the SAT scores of male students with athletics as their only form of extracurricular participation had scores significantly below the national average, whereas athletes with another type of extracurricular participation had scores significantly above the national average. The results suggest that the athlete-only group possessed limited scope in terms of educational aspirations (i.e., more may be going to college to further their athletic involvement), and they were unable to cope with the academic demands at the college level.

EXERCISE AND COGNITIVE FUNCTIONING

At least seventeen reviews have examined the effect of acute (i.e., tested following a single bout of exercise) and chronic (tested following a longitudinal exercise training program) exercise on cognitive function. Some of these reviews concluded that the findings were mixed. However, when these findings were subjected to meta-analysis (Landers & Salazar, 1991), acute and chronic exercise was found to be related to better reaction time (RT) and intelligence (IQ) scores (e.g., Wechsler IQ Test for Children, Iowa Test of Basic Skills, and Goodenough Mental Age). Acute and chronic exercise did not produce any reliable effect on verbal, math, memory, and symbolic reasoning tests. These effects were greatest for subjects that were 16 years of age or less and over age 45.

For tasks emphasizing mental-processing speed, physical fitness seems to be particularly important for the elderly. Bashore (1990) reviewed several behavioral and electrophysiological studies comparing the choice RT performance of young (20–30 years) and old (50–80 years), physically fit and unfit subjects. Compared to the old-unfit subjects, the old-fit subjects performed better and had facilitated information processing. In many cases, the old-fit subjects performed and processed information at approximately the same level as the young-unfit subjects. The young-fit subjects were the best performers. These findings suggest that

physical fitness may help a person sustain a broad range of central nervous system functions, “both those mediating the transmission of primary information and those mediating higher-order cognitive processes” (Bashore, 1990, p. 137).

Black et al. (1990) provided neurobiological evidence for the cognitive effects noted above. Among young-adult female laboratory rats, aerobic training resulted in a greater density of blood vessels than untrained animals. In addition, rats given acrobatic training (learning to traverse an elevated path consisting of balance beams, seesaws, and rope bridges) had more synapses per Perkinje cell in the cerebellar cortex than exercising or inactive animals. These findings implicate the importance of exercise, particularly aerobic exercise involving some degree of learning, in brain morphological changes relevant to cognitive functioning.

(See also: MULTIPLE INTELLIGENCES THEORY.)

BIBLIOGRAPHY

- BASHORE, T. R. (1990). Age, physical fitness, and mental processing speed. In M. P. Lawton (Ed.), *Annual review of gerontology and geriatrics: Vol. 9. Clinical and Applied Gerontology*. New York: Springer.
- BLACK, J. E., ISAACS, K. R., ANDERSON, B. J., ALCANTARA, A. A., & GREENOUGH, W. T. (1990). Learning causes synaptogenesis, whereas motor activity causes angiogenesis, in cerebellar cortex of adult rats. *Proceedings of the National Academy of Science*, 87, 5568–5572.
- HARDMAN, K. (1973). A dual approach to the study of performance in sport. In H. T. A. Whiting, K. Hardman, L. B. Hendry, & M. G. Jones (Eds.), *Personality and performance in physical education and sport*. London: Henry Kimpton.
- LANDERS, D. M., FELTZ, D. L., OBERMEIER, G. E., & BROUSE, T. R. (1978). Socialization via interscholastic athletics: Its effects on educational attainment. *Research Quarterly*, 49, 475–483.
- LANDERS, D. M., & SALAZAR, W. (1991). *Effects of exercise on cognitive functioning: A meta-analysis*. Paper presented at the North American Society for the Psychology of Sport and Physical Activity, Asilomar, CA.
- SNYDER, E. E., & SPREITZER, E. A. (1983). *Social aspects of sport* (2nd ed.). Englewood Cliffs, NJ: Prentice-Hall.
- SPADY, W. G. (1970). Lament for the letterman: Effects of peer status and extracurricular activities on goals and achievement. *American Journal of Sociology*, 75, 680–702.
- DANIEL M. LANDERS

ATTENTION *Attention* is a cognitive mechanism that selects relevant environmental information and controls task performance in support of an individual's goals. Although attention often connotes “consciousness” or “effort,” most modern approaches to understanding attention emphasize its role in perceptual selection and cognitive control (see Allport, 1989, for a comprehensive review). Attention has been a focus of serious investigation within psychology and cognitive science for over a century, since William James said:

Everyone knows what attention is. It is the taking possession by the mind, in clear and vivid form, of one out of what seem several simultaneously possible objects or trains of thought. Focalization, concentration, of consciousness are of its essence. It implies withdrawal from some things in order to deal effectively with others . . . [James, 1890, pp. 403–404].

These remain the dominant themes in attention research today.

Cognitive theories of intelligence often include attention as a fundamental component in planning and executing intelligent behavior (e.g., Sternberg, 1985). Attention has two major aspects. First, as James noted, attention is a mechanism for selecting perceptual information from the environment. The ability to efficiently select relevant information and to ignore irrelevant information is crucial in all aspects of intelligent behavior, from classroom learning and complex problem solving to driving a vehicle and hitting a baseball. Second, attention is a mechanism for controlling the performance of two or more simultaneous tasks. For example, a helicopter pilot must maintain control over the aircraft and simultaneously communicate with the control tower; attention is required to govern potential conflicts between tasks and to implement moment-to-moment priorities. These aspects of attention are discussed below, followed by a review of individual differences in attention and how they may contribute to individual differences in intelligence.

PERCEPTUAL SELECTION

Imagine you are at a crowded party, immersed in the sounds of a dozen simultaneous conversations. Even with the noise, you are able to carry on your own conversation with a friend. In so doing, you find you are unable to follow any other conversation without abandoning the one with your friend; the unattended conversations merge into a background babble. However, as your friend is speaking, you suddenly hear your name in a conversation to your left, and your attention involuntarily snaps there. In the process, you completely lose the thread of what your friend had been saying, but you continue to nod and smile as you strain to hear what is being said in the conversation to your left.

This scenario illustrates what is known as the “cocktail party phenomenon,” which inspired the earliest studies of selective attention in the 1950s. This work involved reproducing the multiple conversations of the party in the laboratory by presenting to listeners a pair of messages via headphones, one message in each ear. These early studies confirmed what our thought experiment revealed: it is very difficult to attend to more than one auditory message at a time, but certain aspects of an unattended message (e.g., the listener’s name) will be heard.

Investigators later turned to visual laboratory tasks to study selective attention, in part because by their very nature, auditory messages must be delivered sequentially in time. In contrast, visual tasks permit the simultaneous and very brief presentation of large amounts of information, providing a way to examine in detail the limits of perceptual selection.

Mechanisms of Visual Selective Attention. Selection of visual information from a crowded display can be guided in either of two ways. First, the observer’s goals, intentions, knowledge, and expectations about upcoming events can influence what is selected; this is known as *top-down* control over selection. Second, certain aspects of the visual scene may capture attention independently of what the observer may know or expect or intend; this is known as *bottom-up* control over selection.

The cocktail party phenomenon illustrates both types. Your effort to follow your friend’s conversation (and then to follow the other, “interesting” conversa-

tion and ignore your friend) is an instance of top-down control over selection. When the arrival of your name captures your attention involuntarily, a form of bottom-up attentional control is responsible.

In the laboratory, psychologists have studied both types of attentional control using a task called *visual search*. In the task, the observer is asked to search for an element in a visual scene (for example, the letter *E* within an array of other letters) and to press a button when the letter is found. This is a laboratory version of the everyday task of searching for a specific box of cereal on a grocery shelf or looking for your pencil on a cluttered desk. The time elapsing from the appearance of the visual array until the observer presses the button, called the *response time*, is measured precisely. Response times typically increase systematically as the number of elements in the display increases (just as the time required to find your missing pencil will depend on whether your desktop is otherwise empty or cluttered). This result suggests that observers must select elements in the visual display one at a time until the target is identified; the more elements there are to select, the longer this process will take, with each additional element adding a constant amount of time to the search.

The visual search task can reveal the influence of both bottom-up and top-down control over attention. For example, if the observer is correctly informed that one location in the visual array is more likely to contain the target than any other location, then observers can and do direct their attention to that location in anticipation of the target event; this is an instance of top-down control (e.g., Downing & Pinker, 1985; Posner, 1982). The redirection of attention can be either *overt* (moving the eyes to the high-priority location) or *covert* (directing attention there without moving the eyes from some anchor point, such as a fixation point in the center of the display).

Now imagine that the display contains one element that exhibits some unique attribute (e.g., one that moves or flashes). Attention is often captured by such an element even when the observer has no reason to believe that the target will appear there, and response times tend to be faster when the target happens to be there than when it appears elsewhere in the display. This is an instance of bottom-up control of attention.

Theories of Selective Attention. Most theories of selective attention hold that perception consists of the construction of a series of increasingly abstract internal representations of the world, starting with a raw sensory image and finally achieving an internal model of the current local environment including three-dimensional surfaces and fully identified object representations (e.g., Marr, 1982). According to such theories, attention selects aspects of early representations and delivers this information to visual mechanisms that compute more abstract internal representations.

The major issue addressed by theories of attention concerns when, within this representational framework, selective attention operates. Two major positions can be identified. According to *early-selection* theories, selective attention operates on an early, relatively unprocessed representation of the perceptual input. So, for example, selection can be based on simple sensory properties of the stimulus such as its location, color, or size, but not on more complex aspects of the stimulus, like its name or semantic category. In order to name or categorize an object, that object must be selected and its attributes compared to stored memory representations. This class of theories was first proposed by Broadbent (1958) and later developed by others (e.g., Treisman, 1986). These theories attribute the partygoer's inability to comprehend more than one conversation at a time to the early locus of perceptual selection.

An alternative view was offered by Deutsch and Deutsch (1963) and subsequently advanced by others (e.g., Duncan, 1980). According to these *late-selection* theories, every element in the visual field is fully identified and categorized before selection; only when a response must be based on a perceptual representation does selection occur. Such theories provide an explanation of why you detect your name in an unattended conversation: all unattended words are in fact identified, but only those that are important to you capture your attention.

A great deal of evidence has been collected in the last three decades in an effort to determine which theory is correct, and no definitive answer has emerged. Instead, the evidence points to a compromise (cf. Yantis & Johnston, 1990): Under certain circumstances, when it is prudent to do so, observers can suppress

identification of unwanted perceptual information. Under other conditions, when the observer's current goals demand divided attention, observers can permit perceptual information from multiple sources to be identified and categorized before selection occurs. A flexible locus for attentional selection is therefore implicated.

COGNITIVE CONTROL

One hallmark of intelligent behavior is the ability to accomplish two or more tasks at the same time. The circus performer who keeps nineteen plates spinning atop nineteen rods exemplifies a physical sort of multitasking. Attention permits cognitive multitasking. Almost everything we do involves some kind of overlap among multiple simultaneous tasks: driving an automobile while listening to the radio; running while dribbling a basketball; carrying on a conversation while walking with a friend. In all these examples, there is the possibility of interference between the tasks. For example, if the driver encounters an accident and must visually guide the vehicle to avoid obstacles, then the content of the radio news broadcast will likely go unperceived; if an opponent moves suddenly into the basketball player's line of sight, the dribble may go astray; or if the walker is asked to solve increasingly difficult math story problems, she may slow her pace and eventually stop walking altogether.

The ability to coordinate two or more simultaneous tasks has been a focus of experimental psychologists for several decades. Early studies examined performance with continuous tasks like manually tracking a moving target with a pointer while listening for an auditory target. However, it quickly became clear that it is often possible to rapidly switch attention between such tasks so that momentary lapses in attention to one task might not be detected. Therefore, investigators have increasingly employed pairs of discrete tasks for studying multitask performance.

Theories of Human Performance in Simultaneous Tasks. The experimental paradigm that has provided the deepest insights about cognitive multitasking is the so-called *psychological refractory period* or *PRP* task. In a typical PRP experiment, an observer carries out two overlapping tasks (see Figure 1a). For

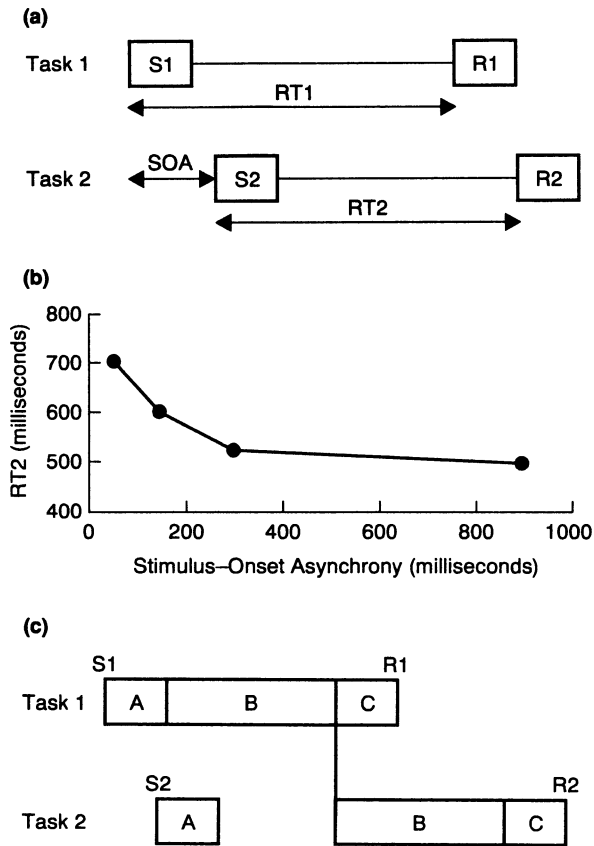


Figure 1
Psychological refractory period (PRP) task. (a) The sequence of events in a typical PRP task. Stimulus 1 (S1) appears, and the time elapsing until Response 1 (R1) is made is RT1. S2 appears after a delay, called the stimulus-onset asynchrony, or SOA, which is the duration between the onsets of S1 and S2. RT2 is the time elapsing between the onset of S2 until R2 is made. (b) RT2 is plotted as a function of the SOA between Task 1 and Task 2. RT2 decreases to an asymptote as the SOA increases, with a slope near -1 at short SOAs. (c) Hypothetical sequence of mental operations required to carry out simultaneous tasks. Boxes indicate duration of each of three successive mental operations: perception (Stage A), response selection (Stage B), and response execution (Stage C). According to the response-selection bottleneck account of the data in panel b, the response selection stage can work on only one task at a time. Task 2 must therefore wait until the response selection stage for Task 1 is complete before response selection can begin in Task 2. Therefore, every millisecond of reduced SOA will add a millisecond to RT2, which yields a slope of -1 at short SOAs in the RT2-SOA function of panel b.

example, Task 1 might require the observer to listen to a tone (called Stimulus 1, or S1), decide whether it is high or low in pitch, and press one key if the tone is high and another if it is low (called Response 1, or R1). Task 2 might require the observer to view an array of digits (called S2), and to determine what the highest digit in the array is and speak that digit into a microphone (called R2). The stimulus-onset asynchrony (SOA) between the two tasks is the amount of time elapsing from the onset of the tone (S1) to the onset of the digit array (S2). If the SOA is quite long, then two tasks will be carried out strictly sequentially (that is, R1 will be completed before S2 appears). As the SOA decreases toward zero (simultaneous onset S1 and S2), the two tasks will increasingly overlap. The experimenter measures the response time for each task (called RT1 and RT2, respectively, defined as amount of time elapsing between S1 and R1 or S2 and R2) as a function of the degree of overlap in the two tasks (as indexed by the magnitude of the SOA). The instructions to the observer are designed to encourage the observer to give highest priority to Task 1, so that any effect of the overlap between tasks will be completely reflected in a slowing of RT2.

Figure 1b shows typical results from such an experiment, with RT2 plotted as a function of SOA. As expected, RT2 is fast if the SOA is long; in this case, Task 2 is carried out in isolation, after Task 1 has been completed. With increasingly shorter SOAs, RT2 becomes increasingly slower. The slope of the SOA versus RT2 function approaches -1 at the shortest SOAs.

In order to understand how observers carry out such overlapping tasks, and in order to explain the pattern of results shown in Figure 1b, attention theorists assume that there is a sequence of mental operations that must be carried out in order to accomplish each task. The sequence of mental operations that is presumed to occur in a PRP task is shown in schematic form in Figure 1c. Each task requires (1) perception of the stimulus, carried out by Stage A in the figure; (2) selection of the appropriate response, carried out by Stage B; and (3) execution of the selected response, carried out by Stage C.

There is substantial evidence that interference between the two tasks takes the form of a *response-selec-*

tion bottleneck (McCann & Johnston, 1992; Pashler, 1989). According to this account, only one response can be selected at a time. If the response-selection stage is currently working on Task 1, then response selection for Task 2 must wait until response selection for Task 1 is complete. Other stages, however, can operate simultaneously on two or more tasks. This means that as the SOA becomes shorter, RT2 must become longer, because response selection cannot begin in Task 2 until it is complete in Task 1. In fact, for every additional millisecond removed from the SOA, a millisecond must be added to RT2. That is reflected in a slope of the SOA versus RT2 function that is near -1 . This provides very good evidence that response selection is an immutable bottleneck in the performance of multiple cognitive tasks.

Studies of multitask performance have shown that certain cognitive operations (e.g., perception and response execution) can be carried out simultaneously, but that response selection is a strict bottleneck in performance: response selection for only one task can be carried out at a time. When an individual must carry out multiple simultaneous tasks, attention provides a mechanism for granting access to the bottlenecks in the system so as to optimize performance.

INDIVIDUAL DIFFERENCES IN ATTENTION

A major goal in the study of intelligence is to understand individual differences in cognitive abilities that contribute to intelligent behavior. Because cognitive theories of intelligence incorporate attention as an important component, we must be concerned with characterizing how individuals differ in their ability to attend to perceptual events or to tasks. Traditionally, this topic is divided into two parts: changes in attention within an individual as he or she develops from birth through adulthood, and differences among adults in their attentional capabilities.

Development of Attention. Any effort to understand the development of some cognitive capacity requires systematic measurement of individual differences in that capacity at different ages. Enns (1990) provides a review of recent advances in our understanding of the development of attention.

These studies have revealed that even infants differ significantly in their ability to attend to perceptual events. How do we know this? The question is not trivial, because infants do not yet talk, and so they are unable to follow instructions or to tell us what they experience. Developmental psychologists have therefore devised clever techniques for measuring attention in infants. Among the most useful of these is the *habituation* technique, in which an infant is shown an object repeatedly and looking time is measured. Typically, looking time decreases with repetition: the infant becomes "bored," or habituates to the object, and looks away.

Using this technique, Bornstein (1990) has found that some infants habituate to the repeated presentation of an object after only a few exposures, whereas others habituate much more slowly or idiosyncratically. Rapid habituation is taken as an indication that an infant is predisposed to explore new or changing aspects of the environment, which has obvious adaptive utility. Bornstein and others have also found that infants who habituate rapidly in the first half-year of life tend to score better on various measures of intelligence in the second year of life and beyond. For example, efficient habituators tend to prefer complex objects to simple ones, to perform picture matching and oddity detection very rapidly and accurately, and to explore their environment more comprehensively. These findings are consistent with the claim that attentive ability is an important component of intelligent behavior.

Other investigators have found that as children develop from about age 4 to adulthood, their ability to attend selectively to perceptual information systematically improves. These developmental improvements have been attributed to at least two sources: first, children become better able to make perceptual discriminations as they grow, and second, children's strategic capacity to deploy attention efficiently improves with age. For example, Kaye and Ruskin (1990) found that younger children (ages 7 to 9) were not as able as older children (ages 10 to 12) or adults to allocate attention to visual events according to their known probabilities of occurrence. This appears to be an improvement in a strategic ability (as opposed to a perceptual one).

Individual differences in any cognitive ability are generally investigated by observing the correlation across individuals in the performance of two or more tasks. For example, Lansman, Poltrock, and Hunt (1983) had eighty-five adults perform target detection tasks in the visual and auditory modalities under conditions of focused attention (the target always appeared in a known location—or ear—among nontargets) or divided attention (the target could appear in either of two locations—or ears—among nontargets). They sought to determine whether there is an underlying ability to focus or divide attention that might account for individual differences in both the visual and auditory domains. If such an ability exists, then individuals who performed well on a visual attention task should also perform well on an auditory version of that task, and those who performed poorly on the visual task should also perform poorly on the auditory task. Lansman and colleagues (1983) found that although observers who performed well on one visual task also performed well on the other visual task, there was not a strong correlation between performance on the visual task and performance on the auditory task. This result suggests that there may not be a single, underlying ability to focus or divide attention that determines each individual's relative performance on focused and divided attention tasks.

In contrast to this failure to find a single underlying ability to focus and divide attention, Tipper and Baylis (1987) found a significant correlation between performance on the Cognitive Failures Questionnaire (CFQ), and the ability to selectively attend. The CFQ, which was devised by Broadbent and colleagues (1982), consists of questions like "Do you fail to notice signposts on the road?" Individuals with high scores tend to be more distractible and absent-minded than those with low scores. Tipper and Baylis (1987) asked a group of subjects to complete the CFQ and then to perform a selective attention task. They found that subjects who were unable to effectively focus attention tended to have high scores on the CFQ (indicating everyday distractibility). This finding suggests that there may indeed be a general ability to focus attention, and this ability is reflected in performance on specific laboratory tests of selective attention as well as in performance in everyday tasks like driving or reading.

At the present time, the study of individual differences in selective attention is just beginning; although much is known about the varieties of attention within the individual, there has been little systematic study of attentional differences across individuals. This is an area where great progress can be expected in the years ahead.

CONCLUSION

Attention is a mechanism used to *select* relevant perceptual information from the environment and deliver it to working memory. Attention also *controls* the performance of simultaneous cognitive operations, providing access to those parts of the system that represent multitasking bottlenecks that must be utilized according to current cognitive priorities. Finally, we can expect significant advances in the study of how individual differences in attention may account for individual differences in intelligence.

(See also: PERCEPTION.)

BIBLIOGRAPHY

- ALLPORT, A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 631–682). Cambridge, MA: MIT Press. This chapter provides a comprehensive overview of recent advances in attention research.
- BORNSTEIN, M. H. (1990). Attention in infancy and the prediction of cognitive capacities in childhood. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 3–21). New York: North-Holland.
- BROADBENT, D. E. (1958). *Perception and communication*. Oxford: Pergamon Press. Broadbent's classic book is credited as being the first modern treatment of attention.
- BROADBENT, D. E., COOPER, P. F., FITZGERALD, P., & PARKES, K. R. (1982). The Cognitive Failures Questionnaire (CFQ) and its correlates. *British Journal of Clinical Psychology*, *21*, 1–16.
- DEUTSCH, J. A., & DEUTSCH, E. (1963). Attention: Some theoretical considerations. *Psychological Review*, *70*, 80–90.
- DOWNING, C. J., & PINKER, S. (1985). The spatial structure of attention. In M. I. Posner & O. S. M. Marin (Eds.), *Attention and performance XI* (pp. 171–187). Hillsdale, NJ: Erlbaum.

- DUNCAN, J. (1980). The locus of interference in the perception of simultaneous stimuli. *Psychological Review*, 87, 272–300.
- ENNS, J. T. (Ed.). (1990). *The development of attention: Research and theory*. New York: North-Holland. This is an invaluable resource covering the development of attention throughout the life-span and attentional deficits in special populations.
- JAMES, W. (1890). *Principles of psychology*. New York: Henry Holt. Chapter 10 of this classic work deals with attention, and it endures as fresh and exciting reading.
- KAYE, D. B., & RUSKIN, E. M. (1990). The development of attentional control mechanisms. In J. T. Enns (Ed.), *The development of attention: Research and theory* (pp. 227–244). New York: North-Holland.
- LANSMAN, M., POLTROCK, S. E., & HUNT, E. (1983). Individual differences in the ability to focus and divide attention. *Intelligence*, 7, 299–312.
- MARR, D. (1982). *Vision*. San Francisco: W. H. Freeman. Marr's computational approach to vision has been exceptionally influential.
- MCCANN, R. S., & JOHNSTON, J. C. (1992). Locus of the single-channel bottleneck in dual-task interference. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 471–484. This article provides an exceptionally clear explanation of an important experimental approach to analyzing dual-task performance.
- PASHLER, H. (1989). Dissociations and dependencies between speed and accuracy: Evidence for a two-component theory of divided attention in simple tasks. *Cognitive Psychology*, 21, 469–514.
- POSNER, M. I. (1982). Orienting of attention. *Quarterly Journal of Experimental Psychology*, 32, 3–25.
- STERNBERG, R. J. (1985). Cognitive approaches to intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence* (pp. 59–118). New York: Wiley.
- TIPPER, S. P., & BAYLIS, G. C. (1987). Individual differences in selective attention: The relation of priming and interference to cognitive failure. *Personality and Individual Differences*, 8, 667–675.
- TREISMAN, A. M. (1986). Features and objects in visual processing. *Scientific American*, 254, 114–124.
- YANTIS, S., & JOHNSTON, J. C. (1990). On the locus of visual selection: Evidence from focused attention tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 135–149.

STEVEN YANTIS

ATTENTION DEFICIT HYPERACTIVITY DISORDER Each year, hundreds of children and adolescents are taken to doctors and mental health professionals because of problems in school and at home with behavior and with learning. Many tend toward excessive physical activity, have trouble paying attention, and often seem to act without thinking. Compared to most young people their age, they have trouble finishing things that they start, cannot sit still or stay long in their seats at school, and always seem to be “on the go.” Though the children may not be very aware of these problems themselves, people in their environment, including parents, teachers, friends, and relatives, are often concerned that these behaviors are interfering with the young person's learning at school and their ability to get along with others. *Attention deficit hyperactivity disorder* (ADHD) is the diagnostic term often applied to such individuals. The root causes of this disorder are still largely unknown, though genetic, prenatal, biochemical, and familial abnormalities are often suspected. Studies have shown that half or more of children with ADHD continue to struggle with symptoms of the disorder as adults. Treatments in the form of medications, psychological and family therapy, and special educational programs have been developed to reduce the effects of the symptoms, but no cure has yet been discovered. These treatments are effective in helping most affected individuals concentrate better, learn to think before they act, and slow down to a more normal pace of living. Research into causes, natural outcomes, and more effective treatment programs continues.

Attention deficit hyperactivity disorder is the contemporary term applied to individuals who meet specific diagnostic criteria defined in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders (DSM)*. Previous diagnostic systems have applied other terms to what is essentially the same disorder. These terms have included *attention-deficit disorder (ADD)*, *hyperactive syndrome*, *hyperkinetic syndrome*, *hyperkinesis*, *minimal brain dysfunction (MBD)*, and *minimal brain damage*. Each diagnostic term has reflected an attempt to describe what, at the time, was believed to be the underlying cause or theoretical core of the disorder. The changes in these terms coincide with the progress of scientific understanding through research that has progressively clarified the essence of

the syndrome. ADHD was originally believed to be a disorder of childhood, but research has confirmed that for many it may be a lifelong problem, the nature of which changes in the course of the developmental stages of life.

CHARACTERISTICS OF ADHD

Attention deficit hyperactivity disorder is one of the more common behavioral syndromes affecting children and adolescents. A syndrome is a combination of symptoms that either result from a single cause or occur together so commonly that their joint appearance indicates the presence of some underlying condition. While the characteristics of these disorders have been thoroughly studied for many years, the task of describing them is in some ways a difficult one. Many of the characteristics are present in all children to some degree at particular times. P. Wender (1987) cautions that it is not the symptoms themselves that set children with these disorders apart but the intensity, persistence, and patterning of the symptoms. Accurate diagnosis of such disorders can be made only by a clinician who has evaluated many children with a range of behavioral and emotional problems. As with any syndrome, not all of the characteristics of ADHD are observed in every person with the disorder. Males are about three times as likely to display the disorder as females.

The central features of the disorder are (1) significant problems with attention; (2) impulsivity; and (3) hyperactivity. At least one of these three major features is usually striking in any affected individual. The degree of these symptoms must be greater than normally expected for the individual's age group. When hyperactivity is not present or prominent, the *DSM* specifies the disorder as "undifferentiated attention-deficit disorder."

Attentional difficulties are currently viewed as the chief difficulty experienced by ADHD individuals. Problems with attention will often be observed as distractibility and a short attention span. Children with such problems seem unable to focus their minds on one thing for more than a brief period before they are distracted and drawn on to something else. Impulsivity may be reflected in the person's appearing to act before thinking and failing to learn from experience.

Schoolwork may habitually be done hastily and in a disorderly, messy fashion. These individuals appear to react too quickly to things, never giving themselves the chance to consider their actions or choices. They may blurt out answers in class, have trouble waiting for their turn in a group task, and frequently interrupt others. Hyperactivity, once regarded as the primary manifestation displayed by ADHD patients, is perhaps the easiest symptom to observe. Affected individuals have difficulty sitting still, constantly fidget or twist in their chairs, tap their fingers, wiggle their feet, and give the impression that they are restless and unable to be still for long. Some talk excessively.

Symptoms tend to wax and wane in their intensity. They often worsen in situations requiring quiet, sustained attention, such as listening to a teacher in school or doing lengthy assignments. Symptoms may seem to disappear when the person is very actively involved in a particular task, such as a conversation, physical recreation, or a video game, or in a new situation, such as a first appointment with a counselor. These core symptoms may be accompanied by over-excitability, a tendency to fail to finish things once started, temper outbursts, a very low ability to deal with frustration, erratic moods, difficulties getting along with others, a tendency to lose things, and underachievement in school or work. Many individuals with ADHD also show signs of learning disabilities, conduct disorder, and oppositional disorder, although the relationship between these disorders remains unclear, with much overlap of characteristics.

ADHD AND INTELLIGENCE

The relationship between ADHD and intelligence is not a simple one. ADHD does not affect intelligence per se or performance on intelligence tests in any specific, predictable way. However, in evaluating the ADHD child, adolescent, or adult, it is often helpful to perform intelligence and academic testing as a means of identifying areas of strength and weakness. There is a good deal of overlap between the incidence of ADHD and specific learning disabilities, particularly disabilities involving reading and mathematical problem solving. Testing can help to assess a student's readiness for mainstream classroom participation and identify those ADHD individuals who may need spe-

cial help in these areas. It is also not uncommon for ADHD individuals to display uneven performance on intelligence subtests. While the overall intelligence quotient (IQ) of ADHD individuals may be within the average range, subtests that are particularly sensitive to the effects of attention and concentration difficulties frequently show deficits. This unevenness in intellectual performance can lead some teachers incorrectly to assume that ADHD students are lazy, uncooperative, or not applying themselves fully to classroom tasks, when in truth their attentional problems are preventing them from doing as well as they could otherwise. Testing often helps parents and teachers correctly to calibrate their expectations of academic achievement.

DEVELOPMENTAL COURSE OF ADHD

The problems displayed by a child with ADHD tend to change with age. Some features disappear altogether, while others change in intensity or form. Wender (1987) reviewed the developmental course of these difficulties in great detail and noted that the problems described below may appear in only some of these developmental stages. In most cases, signs of significant problems are identified by age 7 or 8, though some come to light much earlier or much later. Both the intensity and the duration of these problems are highly variable. If the disorder first appears in infancy, the ADHD child is likely to be much more irritable, fussy, and prone to colic than most infants. Sleep disturbances are common. As toddlers, such children continually "get into things," run and climb constantly, move rapidly from one activity to another, and do not appear to respond to parental discipline. By preschool age, problems with attention and social adjustment are often appearing. The child may appear not to listen and seem not to learn from experience or punishment. At this stage, the child's short attention span, difficulties dealing with frustration, and frequent temper tantrums make learning and smooth social interactions difficult. Problems getting along with other children because of bossiness, teasing, and selfishness create even more chaos.

By the time the child starts school, restlessness and fidgeting are often the most obvious problems. Teachers complain that the child cannot sit still or stay

seated when required to do so. Significant problems with following instructions and completing tasks often surface when the demands of classroom structure are confronted. Early signs of learning difficulties become more apparent, and often worsen, as the child is called upon to learn more and more complex material and work more and more on an independent basis. When a child is 9 or 10 years of age, academic and disciplinary troubles often become more prominent, and problems with lying, cheating, and stealing may emerge. Difficulty with reading and arithmetic are typical, and the child may be criticized for messy and incomplete schoolwork. Coordination problems frequently come to light, and sporting activities are often the source of great disappointment and anxiety. Chores and tasks at home are regularly forgotten or incompletely done. Effects of such academic, behavioral, and social difficulty may seriously impair the development of the child's sense of self-worth and self-esteem. Attention-seeking behaviors, depression, anxiety symptoms, and social isolation often result.

By an ADHD patient's adolescence, patterns of irresponsibility and antisocial behavior may be most notable, while hyperactivity and obvious attentional problems, which often remain, may become less prominent. If symptoms of the disorder have gone unchecked over several years, the youngster may find his or her environment a negative, punishing, frustrating place. By this time, the child may feel chronically out of place, unable to please, and incapable of doing right. Frustration is often expressed outwardly. Truancy, vandalism, fighting, recklessness, running away from home, substance abuse, and other antisocial behaviors often overshadow academic deficiency. Treatment of ADHD that first emerges in adolescence is especially challenging. Rebelliousness, a normal part of adolescence, is more powerfully expressed in late-blooming ADHD cases, making cooperation with the treatment process especially difficult. Children identified and treated before this stage of development appear to have a better prognosis than those first diagnosed in adolescence (Wender, 1987).

Until the late 1970s, most clinicians believed that these problems were usually outgrown as the child's brain matured. Though still controversial, evidence continues to mount that a significant proportion of ADHD children carry some symptoms of the disorder

into adulthood and that treatments effective for children offer substantial help to such adults. Most frequently, major hyperactivity symptoms diminish or disappear by adulthood, although many patients continue to fidget and complain of restlessness. Difficulties with inattention and impulsivity often persist most prominently. Such adults remain easily distracted, fail to finish what they start, appear not to listen well, and have trouble persisting when required to sit and attend for long periods. They have difficulty organizing their work, often act before thinking, shift frequently from one activity to another, and are easily frustrated when things go too slowly or when taking turns. Their concentration problems may be minimal when they focus on material in which they are interested, but may interfere consistently with performance on less interesting tasks. Some continue to find it very difficult to watch television or to read for more than a few minutes. Self-control deficits, including temper problems, tendencies to interrupt others excessively, low stress tolerance, and recklessness, may continue to interfere with their lives. Many complain of frequent mood swings and significant problems in forming and maintaining mature relationships. Long-term follow-up studies suggest that a significant number become actively antisocial and engage in serious criminal activity and serious substance abuse (Klein & Mannuzza, 1991).

POSSIBLE CAUSES

Parenting a child with unrecognized symptoms of ADHD is a frustrating and perplexing experience. By the time the child's behavior has been brought to the attention of psychologists, counselors, or physicians, many angry and disappointing years may have already passed. Parents often feel guilty, helpless, and incompetent. Many blame themselves and naturally ask what the cause is. Despite many years of research, the precise cause of the disorder is still unknown. Nonetheless, scientific research has collected enough information to indicate that the disorder is the result of a dysfunction of the central nervous system and that its causes may be many. This conclusion is based upon the combined results of years of study of the relationship between ADHD and other psychiatric and brain diseases, the genetic patterns of the disorder, biochemical characteristics of those suffering from the disorder,

and the results of treatment studies. Information from each of these lines of reasoning will be briefly reviewed.

Shortly after World War I, there was a worldwide epidemic of von Economo's encephalitis, a viral brain infection that was often fatal. Children who recovered were left with symptoms of hyperactivity, impulsivity, and learning disabilities, the descriptions of which closely match many modern-day observations of ADHD symptoms. Attempts to treat these symptoms led to the discovery of drugs that are successfully used today in the treatment of ADHD, described below. While the vast majority of individuals with ADHD have never had such a brain infection, the similarities in symptoms suggest that certain abnormalities in brain chemistry probably underlie this disorder. It is important to note that current research has shown that most ADHD patients are not brain-damaged, as early theories about the causes of ADHD had suggested. Though brain trauma or disease can leave an individual with ADHD, these are by no means the most common causal factors. ADHD appears most often to result from abnormalities in brain chemistry due principally to genetic factors. Problems in the development of the fetus also show some relationship to the development of ADHD. Such problems can be the result of many factors, including prematurity and maternal alcohol use (Nanson & Hiscock, 1990; Wender, 1987).

A major line of evidence about the causes of ADHD comes from study of family members of ADHD-afflicted individuals. Some psychiatric disorders show a pattern wherein a problem is passed from one generation to the next, indicating a genetic component to the disorder. Most research on the incidence of ADHD in the general public suggests that it occurs in 5–7 percent of all people and 6–9 percent of school-age children. If ADHD has such a genetic component, relatives of individuals with ADHD would be more likely than the general population to display ADHD symptoms themselves. This is exactly what has been found. Studies by J. Biederman and colleagues (1990, 1992) show that 16–25 percent of relatives of ADHD-diagnosed persons also display the disorder, suggesting that relatives are two to four times more likely to have the disorder than members of the general population. Furthermore, up to 30 percent of parents of ADHD chil-

dren also display significant signs of the disorder (Biederman et al., 1986). For a number of years, popular literature proposed that ADHD resulted from certain food allergies or sugar intolerance, but carefully controlled scientific research has failed to support these theories (Zametkin, 1989).

TREATMENT

Treatment strategies vary according to the severity of the disorder and the age of the patient. Individuals with relatively mild symptoms are often able to overcome their difficulties through some combination of counseling, specialized educational help, and parental training. More-seriously affected patients often require medications in combination with more-specialized psychological and educational help.

Amphetamines and related compounds that stimulate the central nervous system have the primary effect of activating the nervous system and increasing wakefulness, decreasing fatigue, and inducing euphoria. In higher doses, these drugs produce agitation, hyperactivity, and insomnia. When abused habitually or used in very high doses, they can produce paranoia and psychosis. In the late 1930s it was accidentally discovered that these compounds had a "paradoxical," or reverse, effect on patients who today would probably be described as ADHD. Rather than further stimulating hyperactive, impulsive, inattentive children, many of whom had mild to moderate forms of brain damage, these medications helped them attend better, reduced their level of physical activity, and appeared to reduce their impulsivity. The value of these medications was quickly explored and established. In ADHD children they were found to have a positive impact on attention, impulsive behavior, hyperactivity, and mood. This led some clinicians to propose that such a positive response in individuals with no history of brain trauma was an indication that they suffered from a form of minimal brain damage or dysfunction that displayed itself as symptoms of what is now called ADHD. Research has helped to clarify this paradoxical effect. It was found that low doses of stimulant drugs had an effect on normal men and boys similar to that observed in those diagnosed as hyperactive. These compounds decreased motor activity, increased attentional focus and vigilance, and improved learning abilities in

normal people in much the same way as in hyperactive individuals (Rapoport et al., 1980). Thus, what was initially thought to be an unusual effect of stimulant medications proved to be a more natural, dose-related effect.

Stimulant medications continue to be the predominant treatment for these disorders. The most commonly used stimulants include Ritalin, Dexedrine, and Cylert. In one survey of primary-care doctors, 88 percent of children considered hyperactive had received Ritalin (Wolraich et al., 1990). In another survey, D. Safer and J. Krager (1988) reported that the percentage of elementary schoolchildren receiving medication treatment for hyperactivity in Baltimore County, Maryland, rose from 1 percent in 1967 to just below 6 percent in 1987. The popularity of this treatment approach reflects its effectiveness in significantly helping up to 80 percent of diagnosed ADHD individuals and the relative rarity of serious side effects. Antidepressant medications are often helpful in those individuals not able to benefit from stimulants (Greenhill, 1989).

Treatment of this disorder is not a matter of simply prescribing the correct medication and waiting to see if it works. While the medication clearly helps to reduce the core symptoms of inattention, impulsivity, and hyperactivity, treatment of the social, academic, and emotional problems that remain requires the work of an integrated team of care providers. Successful treatment of children and adolescents with ADHD usually results from the coordinated efforts of parents, family members, teachers, a physician, and one or more other mental health professionals. Depending on the nature and degree of the particular child's disorder, psychologists, school counselors, social workers, and educational specialists may need to be involved. These professionals help patients and their families develop effective learning methods and communication patterns, more-potent disciplinary strategies and parenting methods, and better techniques of self-control for the individual. With an adult patient, it is almost mandatory to involve the patient's spouse, and on occasion, it may be helpful to request the help of the individual's employer or supervisor.

Treatment of this multimodal nature has been clearly shown to be effective in the short run in alleviating core symptoms and reducing the effects of ac-

ademic and social deficits (Henker & Whalen, 1989). Questions about the long-term effects of treatment remain, in part because of the difficulty of conducting scientifically well-controlled studies across many years. Research to date has been inconclusive, and attempts to clarify which sorts of individuals benefit most from which forms or combinations of treatments continue.

CONCLUSION

Decades of research cannot be summarized adequately in an entry of this length. Attention-deficit hyperactivity disorder is a psychiatric condition whose causes, course, treatments, and label continue to evolve. Controversy as to the nature of the disorder, its relationship to other psychiatric conditions, and the best treatment approaches continues in the scientific literature. Many questions remain. Answers to these questions appear to have far-reaching implications in areas as diverse as education, neurobiology, parenting, substance abuse, violence, and criminality. Many years will probably pass before all of the answers are in.

(See also: HYPERACTIVITY.)

BIBLIOGRAPHY

- AMERICAN PSYCHIATRIC ASSOCIATION. (1987). *Diagnostic and statistical manual of mental disorders* (3rd, rev. ed.). Washington, DC: Author.
- BIEDERMAN, J., FARAONE, S., KEENAN, K., KNEE, D., & TSUANG, M. (1990). Family-genetic and psychosocial risk factors in DSM-III attention deficit disorder. *Journal of the American Academy of Child and Adolescent Psychiatry*, 29, 526-533.
- BIEDERMAN, J., FARAONE, S., KEENAN, K., BENJAMIN, J., KRIFCHER, B., MOORE, C., SPRICH-BUCKMINSTER, S., UGAGLIA, K., JELLINEK, M., STEINGARD, R., SPENCER, T., NORMAN, D., KOLODNY, R., KRAUS, I., PERRIN, J., KELLER, M., & TSUANG, M. (1992). Further evidence of family-genetic risk factors in attention deficit hyperactivity disorder. *Archives of General Psychiatry*, 49, 728-738.
- BIEDERMAN, J., MUNIR, K., KNEE, D., HABELOW, W., ARMENTANO, M., AUTOR, S., HOGE, S., & WATERNAX, C. (1986). A family study of patients with attention deficit disorder and normal controls. *Journal of Psychiatric Research*, 20, 263-274.
- GREENHILL, L. (1989). Treatment issues in children with attention-deficit hyperactivity disorder. *Psychiatric Annals*, 19, 604-613.
- HENKER, B., & WHALEN, C. (1989). Hyperactivity and attention deficits. *American Psychologist*, 44, 216-223.
- KLEIN, R., & MANNUZZA, S. (1991). Long-term outcome of hyperactive children: A review. *Journal of the American Academy of Child and Adolescent Psychiatry*, 30, 383-387.
- NANSON, J., & HISCOCK, M. (1990). Attention deficits in children exposed to alcohol prenatally. *Alcoholism*, 14, 656-661.
- RAPOPORT, J., BUCHSBAUM, M., WEINGARTNER, H., ZAHN, T., LUDLOW, C., & MIKKELSEN, E. (1980). Dextroamphetamine: Cognitive and behavioral effects in normal and hyperactive boys and normal adult males. *Archives of General Psychiatry*, 37, 933-946.
- SAFER, D., & KRAGER, J. (1988). A survey of medication treatment for hyperactive/inattentive students. *Journal of the American Medical Association*, 260, 2256-2258.
- WENDER, P. (1987). *The hyperactive child, adolescent, and adult*. New York: Oxford University Press.
- WOLRAICH, M., LINDGREN, S., STROMQUIST, A., MILICH, R., DAVIS, C., & WATSON, D. (1990). Stimulant medication use by primary-care physicians in the treatment of attention-deficit hyperactivity disorder. *Pediatrics*, 86, 95-101.
- ZAMETKIN, A. (1989). The neurobiology of attention-deficit hyperactivity disorder: A synopsis. *Psychiatric Annals*, 19, 584-586.

MARK F. WARD

AUDITORY ABILITIES In the broadest sense, auditory abilities are cognitive abilities that depend on sound as input and on the functioning of our hearing apparatus. The term encompasses simple sensory processes as well as the abilities required for the solution of complex problems that are presented verbally or by musical means. It also refers to the processes that are characteristically measured through the sense of hearing even though other senses may be used for the same purpose (e.g., rhythm, time-dependent and sequentially presented problem-solving tasks).

Prior to being integrated within the theory of fluid (*Gf*) and crystallized (*Gc*) intelligence (see Horn & Stankov, 1982; Stankov & Horn, 1980), the study of

auditory abilities was largely motivated by practical needs, for example, the selection of musicians and of military personnel such as sonarmen and radio operators. This early research falls into two broad categories: studies of *musical abilities* and studies of *speech perception*. The main aim was to identify independent auditory abilities irrespective of their relationship to the abilities from other sensory domains. Factor analysis was the main statistical procedure used for this purpose.

Auditory abilities form three hierarchically organized layers that parallel those from the visual domain. The layers conform to the classical view of cognition, which distinguishes between sensory, perceptual, and higher-order thinking processes. Examples of the tests used to measure some of these abilities are provided below.

At the lowest level are the sensory detection abilities, including simple processes of time estimation, frequency and energy discrimination, spatial localization and binaural hearing and so on. Different sensory detection abilities share relatively little variance among themselves, and although they are obviously involved with the processes of the other two groups, correlations with the abilities from these higher-order groups are, at best, medium in size. Their correlation with abilities from other sensory modalities is essentially zero. At the highest level are the abilities of listening comprehension, temporal tracking and integration, and so on, those which are clearly intellectual in nature in the sense that they depend on grasping complex and demanding relations among elements of the problems. These show correlations with nonauditory intellectual tasks that are of the same order of magnitude as, say, intercorrelations among the all-visual *Gf* and *Gc* tasks. While they qualify as auditory in the broad sense of the word, for many of these abilities the reliance on sound is incidental; the same processes could be adequately measured with other sensory inputs.

Auditory abilities in a narrow sense involve perceptual processes intermediate between the two extremes of sensory detection and dealing with complex relations. Within the visual domain their counterparts are broad spatial visualization (*Gv*) abilities. Since the original work on musical abilities carried out by Seashore during the second decade of the twentieth century,

many different types of auditory tests have been developed. The most important primary auditory abilities within this array of tests are (a) the ability to perceive speech under distortion/distraction (SPUD); (b) the ability to discriminate among tonal sound patterns (DASP); and (c) the ability to maintain and judge rhythmic sequences of sounds (MaJR). These three primary abilities, together with parts of lower- and higher-level measures that also use sounds as stimuli, define broad auditory factor (*Ga*) at the second order of analysis. This factor is independent of the remaining broad factors of the *Gf/Gc* theory.

In addition to these positive findings, there are uncertainties about the existence of some primary auditory abilities. For example, research on individual differences has been equivocal about the ability to discriminate sounds with respect to loudness. Also, fine-grained studies of listening comprehension indicate that a separate ability to process phonological aspects of speech—that is, discrimination among the phonemes—may exist. It is not known whether, within a broad sample of cognitive tasks, this latter ability would be absorbed by the SPUD factor or whether it would remain on its own. Finally, relative to the processes of visualization, individual differences within the auditory domain are poorly explored. Late-twentieth-century technological advances make the work in this area easier and may lead to a discovery of new factors.

What influences lead to the emergence of a broad auditory ability that make it stand apart from the other cognitive abilities? Lazar Stankov (1983) proposed that the most important aspect of auditory processing is “competition.” This is conceptualized in a very broad sense to encompass a number of distinct influences. In the case of MaJR, internal metabolic processes give rise to a sense of time that competes with the externally introduced tempo and rhythmic patterns. Individual differences in judgments of rhythm and tempo are dependent on the competition between these internal and external processes. In the case of SPUD, the competition is between the signal (speech) and noise (distraction)—individual differences arise because of different sensitivities to the combined effects of these two sources. Finally, in the case of the DASP ability, competition between the signals that originate

from the two ears—for example, the occlusion of the ipsilateral by the contralateral pathway—may be a critical factor. These three types of competition—clearly the result of an interaction between genetic and environmental influences—may share a commonality sufficient to lead to the emergence of the *Ga* factor at the second order. While some kind of competition may be a part of nonauditory broad ability factors, their emergence is likely to be affected by individual differences in some additional aspects of neural organization. For example, individual differences in the organization and localization of processes within the right hemisphere (for broad visualization, *Gv*) and left hemisphere (for *Gc*) may provide a basis for the separation of abilities within these factors. There is little evidence to suggest that *Ga*, like *Gf*, is localized within a particular hemisphere of the brain, although some primary abilities of *Ga* may be so localized. For example, SPUD appears to be lateralized to the right and DASP largely to the left hemisphere.

In addition to charting the poorly explored regions of cognition, the study of auditory abilities has facilitated an easy translation of some of the theoretically interesting tasks developed by the experimental cognitive psychologists into a format suitable for research in individual differences. Thus, many measures of attentional processes (e.g., sustained attention, search, attention switching) are based on auditory modality. Particularly important is the work on divided attention that employed competing tasks. These are the tasks in which two measures of intelligence (e.g., the Letter Series Test and Tonal Reordering Test) are given simultaneously, one through the left and another through the right ear. The results of this work have been interpreted in terms of the available pool of processing resources and a central construct of intelligence—Charles Spearman's notion of mental energy.

HOW ARE AUDITORY ABILITIES MEASURED?

Given the three-layered structure of auditory abilities, it is necessary to describe tests that are illustrative of each of these layers.

Measures of Sensory Detection. The most commonly employed of these types of measures are

the hearing acuity assessments provided by the typical audiograms. They contain information about the hearing thresholds for single pure tones of varying frequencies. Most research of the late twentieth century has focused on the following three rather similar tests, which involve comparisons between pairs of tones.

1. Seashore's test of pitch differences. This is a fifty-item test consisting of pairs of tones differing in frequency by between 2 hertz and 17 hertz. Each tone lasts longer than 500 milliseconds. All sound frequencies are well within the normal adult hearing range. The test begins with easy items and progresses to more difficult items; the task is to state if the two tones are the same or different. The measure is the number of correctly answered items.

2. Frequency discrimination for short tones (Raz, Willerman, & Yama, 1987). This test involves hearing two tones, each for 20 milliseconds duration and with a silent gap lasting 850 milliseconds between them. The peak signal sound pressure level at the headphones is a clearly audible 96 decibels. Pairs of tones differ in frequency; the task is to state on every trial whether the first or the second tone has higher pitch. At the beginning of the trial, the tones are 870 hertz and 770 hertz. The frequency difference is gradually reduced following the accepted psychophysical procedures for the determination of a differential threshold. The measure is the threshold at which difference is detected.

3. Auditory inspection time (Deary, 1992). The stimuli to be discriminated are two square wave tones (870 Hz and 770 Hz) played at the 80 decibel sound pressure level. Both tones last for equal time periods. The length of the time interval varies between pairs, however. The aim is to establish the shortest time interval—that is, the duration of tones—for which the subject can (with 90% accuracy) state whether the order of the tones within a pair was "high–low" or "low–high." As found in the visual inspection time experiments, the measurement of auditory inspection time depends on the control of certain conditions. Thus, there is no gap between the playing of the tones within a pair, and there is masking stimulus immediately after the second tone. The measure is the shortest tone duration for correct answers.

Measures of Perceptual Processes. These are examples of measures for the three primary abilities of *Ga*.

1. Expanded speech (a measure of SPUD). Short sentences are played at a slower speed than that used in recording. The task is to write the sentence. The measure is the accuracy of the sentences written over ten to twelve sentences.

2. Chord decomposition (a measure of DASP). A three-note chord is followed by four answer choices in which three notes are played separately. The task is to indicate which answer choice involves the same three notes as were in the original chord. The measure is the number of correct answer choices over fifteen to twenty items.

3. Tempo (a measure of MaJR). The task is to continue to count a beat established by a metronome and after varying lengths of time write the number to which the beat has been carried. The measure is the accuracy of the count.

Measures of Intellectual Processes. The following three tests are clearly more demanding than those mentioned above.

1. Cloze. The task is to write two words that are missing in ten to fifteen spoken sentences of eight words each. Missing words are indicated by clicks. This test is a measure of the auditory listening comprehension primary ability and *Gc* at the second order.

2. Tonal reordering. Three notes are played, and subjects are asked to "label" these notes with numbers corresponding to the order in which they were presented. After a short pause, the same three notes are played in a different order. The task is to write the number "labels" in an order corresponding to the order of presentation in this second hearing. Example: do ("1"), re ("2"), mi ("3") are played in that order first and then shortly afterward as re, mi, do. Answer: 2, 3, 1. To get reliable measures, fifteen to twenty such tasks are presented. This test measures temporal tracking primary ability and *Gf* at the second order.

3. Triplet numbers. The stimulus material for this test consists of a randomly chosen set of three different digits ("triplets") presented successively, one per second, on the computer screen. These triplets are separated by a four-second interval. The subject has to press the "Yes" key if the first digit is the largest and the second digit is the smallest or if the third digit is

the largest and the first digit is the smallest. Otherwise, the "No" key is to be pressed. The measure is the number correct. This test measures sustained attention primary ability and *Gf* at the second order.

WHAT IS THE RELATION OF AUDITORY ABILITIES TO INTELLIGENCE?

The importance of auditory abilities for a normal development of intelligence has been recognized for a long time. For example, the opening paragraph of Diane Ackerman's chapter on "Hearing" reads as follows. "In Arabic, absurdity is not being able to hear. A 'surd' is a mathematical impossibility, the core of the word 'absurdity,' which we get from the Latin *surdus*, 'deaf or mute'.... The assumption... is that the world will still make sense to someone who is blind or armless or minus a nose. But if you lose your sense of hearing, a crucial thread dissolves and you lose track of life's logic" (see Ackerman, 1991, p. 176). Indeed, in the absence of intensive specialized education for blind and deaf, the blind suffer smaller intellectual damage than the deaf because of sensory deficit. This is perhaps due to the fact that the prime channel for communicating, and therefore for acquiring any knowledge during childhood, is auditory. Deaf children may be particularly affected because of the reduced exposure to abstract concepts. As a consequence, the average blind person's measured intelligence quotient (IQ) tends to be higher than the average IQ of the deaf. Within the normal population of children, one of the best predictors of performance on intelligence tests at age 11 is the listening comprehension test taken four years earlier.

While studies of individual differences in auditory abilities provide clear support for the views consistent with the comprehensive theory of fluid and crystallized intelligence (Horn, 1988), two recent attempts to link these abilities with intelligence focus on issues that have been peripheral to the main thrust of the *Gf/Gc* theory.

First, some investigators who have been searching for the "basic" processes of intelligence have focused on auditory abilities. This approach is closely linked to the late-twentieth-century interest in "biological" sources of individual differences. Basic and simple pro-

cesses, rather than cognitively complex and demanding processes, are seen as being close to biological measures such as electroencephalogram recordings, nerve conduction velocity, metabolic rate for glucose, or even some physical properties, for example, the ratio of white to gray matter in the cortex. Those working within this paradigm expect to find a single psychological process that underlies all intelligent behavior, which can then be linked to these biological measures.

One line of research with auditory abilities points to the importance of sensory acuity (i.e., frequency discrimination); another line places particular emphasis on the role of mental speed (i.e., auditory inspection time). Descriptions of the tasks used in this work are given above. Indeed, significant correlations between frequency discrimination for short tones and intelligence and auditory inspection time and intelligence have been reported, supporting both interpretations. Despite the improvement in the measurement of sensory processes over the situation that existed at the beginning of this century, however, raw correlation between auditory inspection time and intelligence is not higher than 0.30 (Deary, 1992). Raz, Willerman, and Yama (1987) have reported correlations between intelligence and measures of frequency discrimination that are between 0.40 and 0.50 but similar to many other studies interested in the "basic" processes, these correlations were obtained with small samples of subjects. Much of the work on "basic" processes of intelligence has employed simplistic conceptions of intelligence—as measured with a single matrices test. This research thus has neglected to consider correlations with potentially important measures, such as those of perceptual speed and broad auditory function, which mediate correlations between sensory processes and intelligence. Also, we do not know the extent to which processes captured by these new sensory tests of intelligence differ from those captured by Seashore's test of pitch differences. Since it is known that Seashore's test has a relatively low correlation with measures of intelligence, if strong similarity of these processes were to be confirmed, sampling error would be a reasonable explanation for the reported significant correlations with intelligence.

Even though it is premature to claim that this work has added substantially new information to our knowledge about the relationship between sensory auditory

abilities and intelligence, the weak link between simple sensory processes and intelligence has been reaffirmed. There is also an outside chance that the relationship may be stronger than previously thought.

In Howard Gardner's (1983) oligarchic theory, musical intelligence is one of seven multiple intelligences. In the theory of fluid and crystallized intelligence, *Ga* is one of nine intelligences. In Gardner's theory, central aspects of musical intelligence are pitch and rhythm, which correspond to DASP and MajR primary abilities, respectively. Those auditory abilities that deal with the verbal material (i.e., listening as captured by SPUD) belong to what he calls linguistic intelligence. He considers auditory—and oral—aspects of language as central to linguistic intelligence. Gardner points to several lines of evidence in support of his position, namely, potential isolation by brain damage, the existence of idiots savants, prodigies, and other exceptional individuals, a distinctive developmental history, the identifiable set of core operations, and susceptibility for encoding in a symbol system. He notes that factor-analytic findings provide criteria for distinguishing independent intelligences. The evidence Gardner cites has been used to build the theory of fluid and crystallized intelligence. Within the auditory domain, Gardner's theory obviously supports the independence of primary abilities, but he has nothing to say about broad ability organizations. Much of his evidence, however, is new and fresh.

Psychologists have often assumed that higher mental processes that are tapped by the typical tests of intelligence depend relatively little on input modality. Relatedly, it is sometimes assumed that what distinguishes, say, a good musician from an exceptionally good musician, is not auditory ability per se but the presence of a higher IQ in the latter. Perhaps some aspects of our thinking about intelligence need to be changed. The existence of broad perceptual factors such as *Gv* and *Ga* suggest that the gap between the lower-order perceptual and high-order intellectual processes may be small, especially among people who spend a lot of time in activities that depend on the functioning of visual and auditory modalities. It is on this basis that we can follow Gardner's suggestion to talk about "visual" and "auditory" intelligence.

(See also: MUSICAL ABILITY; MUSICAL INTELLIGENCE.)

BIBLIOGRAPHY

- ACKERMAN D. (1991). *A natural history of the senses*. New York: Random House, Vintage Books.
- DEARY, I. J. (1992). *Auditory inspection time and intelligence*. Unpublished doctoral dissertation, University of Edinburgh.
- GARDNER, H. (1983). *Frames of mind: A theory of multiple intelligences*. New York: Basic Books.
- HORN, J. L. (1988). Thinking about human abilities. In J. R. Nesselrode (Ed.), *Handbook of Multivariate Psychology* (pp. 675–685). New York: Academic Press.
- HORN, J. L., & STANKOV, L. (1982). Auditory and visual factors of intelligence. *Intelligence*, 6, 165–185.
- RAZ, N., WILLERMAN, L., & YAMA, M. (1987). On sense and senses: Intelligence and auditory information processing. *Personality and Individual Differences*, 8, 201–210.
- STANKOV, L. (1983). The role of competition in human abilities revealed through auditory tests. *Multivariate Behavioral Research Monographs*, No. 83–1, p. 63.
- STANKOV, L., & HORN, J. L. (1980). Human abilities revealed through auditory tests. *Journal of Educational Psychology*, 72, 1, 19–42.

LAZAR STANKOV

AUTISM The syndrome of infantile autism (also known as *autistic disorder*, *Kanner's syndrome*, or *early infantile autism*) was first described in 1943 by Leo Kanner, who reported on eleven children exhibiting an apparently congenital lack of interest in other people. These children were highly interested, however, in unusual aspects of the inanimate environment. For several decades after Kanner's initial description, research on autism and related conditions suffered from both a lack of consensus on aspects of syndrome definition and assumptions of continuity between these conditions and severe forms of mental illness in adults, particularly schizophrenia (Kolvin, 1971). The presumed link with schizophrenia derived from the similarity in severity of the conditions. The assumption was consistent with the very broad views of the syndrome boundaries of schizophrenia. It was suggested by Kanner's use of the term *autism*, which had previously described the self-centered thinking observed in schizophrenia. Researchers disagreed about the role of ex-

periential factors in the pathogenesis of the autistic syndrome.

Over the past two decades several lines of evidence have supported the existence of autism apart from schizophrenia and have suggested the importance of neurobiological factors in syndrome pathogenesis. For example, rates of schizophrenia do not increase among first-degree relatives of individuals with autism, and the clinical features of children with autism differ substantially from those observed in children with early-onset schizophrenia (Werry, 1992). Individuals with autism do not appear to have markedly increased rates of schizophrenia.

Autism is now grouped within the class of pervasive developmental disorders (PDD) along with similar disorders: Asperger's disorder, Rett's disorder, childhood disintegrative disorder (Heller's syndrome or disintegrative psychosis), and pervasive developmental disorder not otherwise specified (PDD–NOS World Health Organization, 1992). Although the syndrome of autism has been studied extensively and its existence established, the “nonautistic” PDDs have been explored less frequently and their VALIDITY, particularly apart from autism, is more controversial.

CLINICAL FEATURES

Kanner's original report outlined features still viewed as central aspects of autism. These include its very early onset (from birth or within the first three years), disturbed social development, absent or markedly deviant speech, and unusual responses to the environment (e.g., resistance to change or “insistence on sameness,” stereotyped motor movements, idiosyncratic interests or responses to the environment). A diagnosis of autism implies very early onset of deficits in the areas of social and communicative disturbance and unusual environmental responsiveness.

Onset. A series of studies has confirmed that autism develops at or shortly after birth. In most (approximately 75%) cases autism appears to be present from birth or within the first year of life. In a smaller number of cases the early development of the child is reported to be reasonably normal. Delayed case detection may result from lack of parental sophistication or from denial, the degree of any associated mental re-

tardation in the child, and a lack of knowledge among primary health care providers. Information on the nature and pattern of onset is particularly important relative to diagnosis because autism almost never develops after age 3. On the other hand, childhood disintegrative disorder develops after a relatively prolonged and unequivocally normal development.

Social Development. For the normally developing baby, interest in the social environment is marked, even from the first weeks of life as the human face and social interaction appear to be of central interest to the child (Volkmar, 1987). As suggested by Kanner (1943), the child with autism appears to lack this early predisposition toward social interaction. In most cases parents report that the affected child exhibited a marked lack of interest in them or other people from very early in life, although interest in aspects of the inanimate environment was strong. The usual milestones of early social development (e.g., the social smile, simple forms of imitation and reciprocal interaction, and development of differential attachments to parents) are typically delayed severely and deviant when they do develop (Volkmar, 1987). Peer relationships typically fail to develop. Affected individuals have difficulties in use of nonverbal behaviors (e.g., eye contact, facial expression) in the regulation of interaction. Lack of emotional reciprocity is common. Over time certain social skills do develop, although even the highest functioning individuals with autism have difficulties in the integration of social-emotional cues and communication. Social skills relate to developmental level. Younger and lower-functioning individuals usually appear to be the least interested in social interaction, and the highest-functioning individuals often interact in an odd or eccentric fashion (Wing & Gould, 1979).

Communicative Development. Failure of the child to develop language is a common complaint. Parents may worry that the child is deaf, although usually the child will be extremely sensitive to aspects of the inanimate environment. The sound of a vacuum cleaner may elicit apparent panic in a child who does not respond to the voice of his parents. Communicative language fails to develop in approximately 50 percent of cases (Pauls, 1987). When it does develop, language has features such as immediate and remote

echolalia, pronoun reversal, failure to intonate appropriately, and marked deficits in use of language in social contexts.

Unusual Responses to the Environment. Unusual responses to the environment may include an inability to tolerate changes or transitions, unusual or idiosyncratic interests or attachments, and stereotyped movements. Purposeless and repetitive, stereotyped movements appear to be preferred modes of activity for the child. These movements may include hand flapping, walking on toes, body rocking, or more complex movements of the body. Capacities for symbolic and imaginative play are impaired; recreational materials may be used in unusual ways, as the child becomes preoccupied with one aspect of an object.

DIAGNOSIS OF AUTISM AND RELATED CONDITIONS

The official American and international (World Health Organization, 1992) definitions of autism are conceptually identical. Of earlier definitions, Rutter's (1978) synthesis of Kanner's original work and subsequent research was highly influential. Rutter noted that the deviance in social and communicative development was not attributable to developmental delay alone. It was greater than expected, given associated levels of mental retardation. Certain aspects of the condition—unusual rates or sequences of development, self-injurious behavior, or deviant responses to sensory stimuli—seem to be associated features of the condition. In addition to categorical definitions of autism, researchers have proposed various rating scales, checklists, and other assessments of *dimensions* of function/dysfunction. In some instances the scales are linked explicitly to categorical diagnostic criteria (e.g., Lord et al., 1989). The tremendous range in syndrome expression over developmental level and chronological age is a complication for both categorical and dimensional assessments. The lack of metrics for understanding normal social functioning has been problematic. Assessment instruments that rely on parental report of very early development of the child may not be reliable; on the other hand, teachers or parents may have difficulties with instruments that inquire about highly deviant behaviors. Disagreement around diagnosis is

usually most pronounced regarding individuals at the extremes of syndrome expression—in the profoundly retarded, in the very young, and in individuals with intelligence quotients (IQs) in the normal range. Clearly, the frequency of “autisticlike” behavior increases with decreasing IQ; however, these “autisticlike” behaviors are more likely to include stereotyped movements and unusual responses to the environment.

Nonautistic PDDs. Several conditions resemble autism in one or more important respects. The inclusion of these conditions within the PDD category has been controversial. In Rett’s disorder very early development is normal. In this condition, so far reported only in females, head growth then decelerates, usually in the first months of life, and a loss of purposeful hand movements occurs. Motor involvement is quite striking, and profound mental retardation is typical. Characteristic handwashing stereotypes develop. Although the course and clinical features of this condition are distinctive, confusion with autism may arise, particularly during the preschool years (Tsai, 1992).

In childhood disintegrative disorder (previously known as *Heller’s syndrome* or *disintegrative psychosis*) a relatively prolonged period of unequivocally normal development (usually for several years) is observed prior to a marked loss of skills and development of an “autisticlike” clinical picture. The onset of the condition may be insidious or abrupt; the deterioration in the child’s development is very dramatic. Most definitions require that the early development of the child be unequivocally normal. Researchers once presumed that the condition was uniformly associated with some specific medical condition that might account for the marked regression. This assumption is no longer valid (Volkmar, 1992). Although probably much less common than autism, CDD appears to have an even worse prognosis.

In Asperger’s disorder (also known as *autistic psychopathy*), cognitive and communicative development are within the normal or near normal range in the first years of life. Deficits appear in social interaction and unusual responses to the environment similar to those in autism. Verbal skills are usually an area of relative strength, in contrast to autism. Idiosyncratic interests are common (e.g., in train schedules, snakes, the weather channel). An increased incidence of this con-

dition may occur in family members. The existence of this condition, apart from high-functioning autism, remains the topic of debate (Szatmari, 1992); it is possible that the condition overlaps at least in part with some forms of learning disability, such as the syndrome of nonverbal learning disability (Rourke, 1989). Some studies suggest that individuals with this condition may be at greater risk for development of psychosis in adolescence or adulthood.

The definition of “subthreshold” PDD (PDD–NOS) is essentially a negative one. It is a disorder with some features suggesting but not fully meeting the criteria of autism. Although this condition is probably relatively common, research is highly limited.

Other Disorders. In mental retardation not associated with autism or other pervasive developmental disorder, the pattern of development delay tends to be relatively even. Social and communicative skills are about as delayed as cognitive skills. In the developmental language disorders, delays and deviance are most noteworthy with respect to language-related functions; social and nonverbal cognitive abilities tend to be preserved. In some instances the child with a marked language disorder may develop a rich repertoire of nonverbal communicative skills. In schizophrenia beginning in childhood the pattern of disturbance is similar to that starting in adolescence and adulthood: The clinical picture is characterized by hallucinations, delusions, disturbances in thinking, and so forth (Werry, 1992).

EPIDEMIOLOGY AND COURSE

Studies employing reasonably stringent definitions of autism suggest prevalence rates of approximately 2 per 10,000 children. The condition occurs in all social classes and in both developed and underdeveloped countries. Males are affected four or five times more often than females. When females have autism they are, however, more likely to be severely retarded. Females with autism and IQs in the normal range are extremely uncommon. Apart from PDD–NOS the other conditions included in the PDD class are apparently less common than autism.

In autism social and communicative deficits are most striking during the preschool years. Various factors may act to delay case detection and provision of

a remedial program. Such delays are unfortunate because at least some research suggests the importance of early detection and sustained educational and behavioral interventions. During the school years behavior problems may be more difficult to manage. During adolescence a small number of persons with autism improve while others deteriorate. Seizures may develop at any point during childhood and adolescence. The available outcome data suggest that perhaps 1 to 2 percent of affected adults are capable of an independent existence. Another 33 percent are able to achieve some degree of personal self-sufficiency. About 66 percent of affected adults require very high levels of support. Overall intellectual level and the ability to use language for communication by age 5 years are important prognostic signs.

PATHOGENESIS

Certain aspects of Kanner's original description served as false leads for research, particularly in relationship to syndrome pathogenesis. Although Kanner's emphasis on the early (if not congenital) onset of the condition indicated some role of neurobiological factors, other aspects of his description suggested some influence of experience in pathogenesis. His initial report suggested that the condition was not commonly associated with other medical conditions, mental retardation, or other signs of central nervous system pathology. On the other hand, Kanner's observation of high levels of parental educational and occupational attainment and of unusual patterns of parent-child interaction seemed to suggest that deviant caregiving might be involved. However, studies of parents of children with autism did not reveal specific psychopathology in parents nor did they reveal specific deficits in caregiving. Deviant patterns of parent-child interaction also appeared to be just as likely to reflect deviance in the child rather than in the parents. The impression of normal intellectual levels derived from the observation that on certain parts of IQ tests children with autism performed in the normal or near normal range. Although some degree of scatter in intellectual skills commonly occurs, most individuals (probably about 80%) with autism are also mentally retarded. Longitudinal and other studies revealed a very high frequency of signs of central nervous system

disturbance, such as persistence of primitive reflexes, delayed development of hand dominance, and EEG abnormalities, and abnormalities on brain scan. Approximately 25 percent of autistic individuals develop seizure disorders of diverse types and another 25 percent exhibit abnormal electroencephalograms (EEGs) (Golden, 1987). Associated with autism is a large number of medical conditions of diverse types, such as fragile-X syndrome and congenital rubella. Finally, studies controlling for factors that may bias case detection have suggested no unusual frequency of autism among more highly educated or successful parents. Researchers presume that deviance *in the child* likely accounts for impressions of deviant patterns of parent-child interaction.

Research has continued to support the role of neurobiological factors in pathogenesis. No precise pathological mechanism has been identified as yet. Structural deficits in the central nervous system have been postulated at levels ranging from the brainstem to the cortex. However, the few neuropathological studies conducted have not yet revealed specific defects. As a group, individuals with autism have high peripheral levels of serotonin, a neurotransmitter. This finding is not specific to autism, and its significance is unclear. Although the role of genetic factors initially seemed limited, a concordance of autism clearly increases among siblings of affected individuals, particularly identical twins (Pauls, 1987) and siblings are at increased risk for other developmental problems.

Available evidence suggests some degree of central nervous system involvement but associated neurobiological findings vary considerably. Neurobiologic findings can be subtle and interpretation of research is complicated by various factors. Various neurobiological factors appear to act, through one or more mechanisms, to produce the syndrome of autism. The study of closely related syndromes such as childhood disintegrative disorder appears to be particularly important in the identification of pathophysiological mechanisms.

COGNITIVE FUNCTIONING

Early assumptions of normal cognitive potential derived from performance on certain aspects of traditional tests of intelligence; poor performance on other aspects of these tests supposedly reflected lack of

motivation or “negativism” on the part of the subject. Subsequent research has shown that if full-scale tests are used and are appropriate to the individual overall intelligence is typically within the retarded range. Scores on such tests are reasonably stable and predict adult outcome. Strengths are usually observed in visual-perceptual or memory tasks, and areas of weakness typically involve conceptual reasoning and sequencing of information. Significant scatter is usually apparent on intellectual testing. When both verbal and nonverbal or performance IQ scores can be computed, verbal IQ is typically much lower than nonverbal or performance IQ.

“Splinter skills” or islets of unusual ability may be present in individuals with autism, for example, in relation to tasks that focus on rote memory or block design. A small number of individuals with autism may exhibit truly unusual abilities, such as in drawing, in feats of memory, or mathematical or calendar calculation. At times these abilities are remarkable, even in relation to the normal population. Examples are so-called autistic savant skills (Treffert, 1988). The range of such skills usually is quite narrow and is usually dissociated from the child’s other cognitive skills, contributing very little to the child’s overall adaptation and social interaction. Similarly, a child with autism occasionally has unusual abilities to decode written material or has an apparently large receptive vocabulary; however, such skills usually are not readily generalized so that understanding of what is read or use of vocabulary in conversation is highly limited.

The pattern of intellectual skills in individuals with IQ scores within the normal range has been controversial. At least two subgroups within the broader “autistic spectrum” of conditions appear likely; these appear to correspond to “high-functioning autism” and Asperger’s disorder. In Asperger’s disorder verbal skills are often an area of strength, whereas among the more capable individuals with autism (strictly defined) performance or nonverbal skills are typically advanced in relation to verbal abilities. The cognitive development of very young autistic children has been less frequently studied. The available data on aspects on sensorimotor intelligence in this age group are conflicting.

The role of cognitive deficits in the pathogenesis of autism has been controversial. The social disturbances

are central aspects of syndrome definition, but the relationship of social deficits to cognitive and communicative ones remains unclear (Volkmar, 1987). On the one hand, attempts to identify specific pathognomic psychological processes (related to certain deficits in learning, such as overselectivity or in social-perceptual skills such as self-recognition) have often been more parsimoniously explained by overall levels of cognitive functioning. On the other hand, studies of social development in autism do suggest that delays in social skills are much greater than those predicted by overall cognitive abilities (Volkmar et al., 1993). Clearly social disturbances are primary, at least in terms of syndrome definition.

Some researchers have made attempts to account for the social deficits in autism on the basis of a very specific cognitive disability, that is, related to deficits in capacities for attributions of wishes and feelings to others (the “theory of mind” hypothesis) (Baron-Cohen, 1988). Although attractive in many respects this hypothesis does not appear to explain the severity and nature of the social deficit in autism (Klin, Volkmar, & Sparrow, 1992).

ASSESSMENT AND INTERVENTION

The assessment of the child with autism usually requires the efforts of various professionals. Previous experience with autism on the part of the evaluators is extremely helpful. Assessment instruments should be chosen with consideration of the individual’s age and developmental level. To the extent possible, assessments should be based on administration of well-standardized assessment instruments. In some cases modifications of usual assessment procedures is clinically indicated; in such cases it is important to view results obtained with caution. Assessments of adaptive behaviors, communication skills (including preverbal skills in low-functioning individuals) should be obtained. The child should be observed during both structured and unstructured periods. Historical information may suggest specific additional procedures, for example, neurological or genetic testing, and is critical in making a diagnosis.

The severity of the condition has led to the use of many different treatments for individuals with autism. These have included somatic treatments, drug treat-

ments, psychotherapy, behavior modification, educational and special educational intervention, and dietary change. With few exceptions (notably of behavior modification and special education and, to a lesser extent drug treatments) interventions have not been systematically evaluated. The available data suggest that the most effective procedures over the long term are those based on intensive special educational and behavioral intervention. A limited role exists for the use of certain pharmacological agents, but they are not curative. The use of nonproven interventions is contraindicated, particularly when it puts the child at further risk.

(See also: MENTAL DISABILITIES.)

BIBLIOGRAPHY

- BARON-COHEN, S. (1988). Social and pragmatic deficits in autism: Cognitive or affective? *Journal of Autism and Developmental Disorders*, 17, 379–402.
- GOLDEN, G. S. (1987). Neurological functioning. In D. Cohen & A. Donnellan (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 133–147). New York: Wiley.
- KLIN, A., VOLKMAR, F. R., SPARROW, S. (1992). Autistic social dysfunction: Some limitations of the theory of mind hypothesis. *Journal of Child Psychology and Psychiatry*, 33, 861–876.
- KOLVIN, I. (1971). Studies in childhood psychoses, I: Diagnostic criteria and classification. *British Journal of Psychiatry*, 118, 381–384.
- LORD, C., RUTTER, M., GOODE, S., HEEMSBERGEN, J., JORDAN, J., MAWHOOD, L., & SCHOPLER, E. (1989). Autism Diagnostic Observation Schedule: A standardized observation of communicative and social behavior. *Journal of Autism and Developmental Disorders*, 19, 185–212.
- PAUL, R. (1987). Communication. In D. Cohen & A. Donnellan (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 61–84). New York: Wiley.
- PAULS, D. (1987). The familiarity of autism and related disorders: A review of the evidence. In D. Cohen & A. Donnellan (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 192–198). New York: Wiley.
- ROURKE, B. P. (1989). *Nonverbal learning disabilities: The syndrome and the model*. New York: Guilford Press.
- RUTTER, M. (1978). Diagnosis and definition. In M. Rutter & E. Schopler (Eds.), *Autism: A reappraisal of concepts and treatment* (pp. 1–21). New York: Plenum.
- SZATMARI, P. (1992). The validity of autistic spectrum disorders: A literature review. *Journal of Autism and Developmental Disorders*, 22, 583–600.
- TREFFERT, D. A. (1988). The idiot savant: A review of the syndrome. *American Journal of Psychiatry*, 145, 563–668.
- TSAI, L. (1992). Is Rett's syndrome a subtype of pervasive developmental disorder? *Journal of Autism and Developmental Disorders*, 22, 551–562.
- VOLKMAR, F. R. (1987). Social development. In D. Cohen & A. Donnellan (Eds.), *Handbook of autism and pervasive developmental disorders* (pp. 41–60). New York: Wiley.
- VOLKMAR, F. R. (1992). Childhood disintegrative disorder: Issues for DSM-IV. *Journal of Autism and Developmental Disorders*, 22, 625–642.
- VOLKMAR, F. R., CARTER, A., SPARROW, S. S., CICHETTI, D. V. (1993). Quantifying social development in autism. *Journal of the American Academy of Child and Adolescent Psychiatry*, 32, 627–632.
- WERRY, J. (1992). Child and adolescent (early onset) schizophrenia—A review in light of DSM-III-R. *Journal of Autism and Developmental Disorders*, 22, 601–624.
- WING, L., & GOULD, J. (1979). Severe impairments of social interaction and associated abnormalities in children: Epidemiology and classification. *Journal of Autism and Developmental Disorders*, 9, 11–29.
- WORLD HEALTH ORGANIZATION. (1990). *International classification of diseases: 10th edition. Diagnostic Criteria for Research*. Geneva: World Health Organization.

FRED R. VOLKMAR

DONALD J. COHEN

B

BAYLEY, NANCY (1899–) Nancy Bayley, an eminent developmental psychologist, is best known for the mental and motor developmental scales that were inspired by a normative study she conducted in 3-, 4-, and 5-year-old children, published in 1926 (Bayley, 1926). She did this first normative work on children while she was a graduate student working on her master's degree at the University of Washington. Bayley suspended her work on developmental testing for a brief time while she pursued her doctorate at the University of Iowa; during this period she studied fear reactions in children as measured by the psychogalvanic skin response (Bayley, 1928).

After receiving her doctorate, Bayley resumed work on children's tests, and eventually she refined a battery of items that came to be called the Bayley Scales of Mental and Motor Development, now distributed in a revised edition (1993) by The Psychological Corporation. These scales are widely recognized as the most elaborately standardized scales for assessing the development of children up to 2 years of age, although Bayley herself concluded by the peak of her career (1955, 1958, 1968) that infant tests applied to essentially normal children in the first year of life do not relate strongly to their later intelligence quotients (IQ). Assessment of the parental IQ, she demonstrated, provides greater accuracy in predicting the child's eventual IQ.

In the course of her career, Bayley's research touched on topics that are very contemporary, such as the continuity of psychomotor and intellectual development; the effects of a mother's behavior on her children; the interplay of psychological and somatic aspects of androgyny; the continuity from childhood of general intellectual ability and specific talents in gifted adults; infant vocalizations and their relationship to mature intelligence; and the importance of documenting contextual and ecological factors that may interact with deliberately imposed experimental or testing conditions to determine behavioral outcomes, including IQs and other indices of performance derived from standardized procedures.

Bayley's hallmark was her rigor in measuring behavior in young children, which would lead to measures of intelligence that honor both continuities and maturational shifts in style of functioning.

CAREER

Born in Oregon on September 28, 1899, Nancy Bayley did not enter public school until she was 8 years of age. After high school, she intended to become a teacher but was captivated by psychology as taught by E. B. Guthrie at the University of Washington, where she earned undergraduate and master's degrees. After earning the Ph.D. at the University of Iowa, she

taught for a brief time at the University of Wyoming before moving to the University of California at Berkeley, where she was on the faculty for most of her career. In an important ten-year period (1954–1964), she was chief of the section on child development in the laboratory of psychology at the National Institute of Mental Health in Bethesda, Maryland. There she refined her Mental and Motor Scales as they were used in a National Collaborative Perinatal Project, a study of 50,000 births that explored the origins of cerebral palsy, mental retardation, and other neurological and psychological deficits. At one period in her career (1939–1951), Bayley held concurrent research appointments in psychology and anatomy at Stanford University and at Berkeley.

Nancy Bayley received many awards and honors during the course of her career, including the G. Stanley Hall Award from the Division of Developmental Psychology of the American Psychological Association (APA), the APA Distinguished Scientific Contributions Award, and the Gold Medal Award of the American Psychological Foundation. Bayley was a fellow of both the APA and the American Association for the Advancement of Science.

Bayley was President of the Division of Developmental Psychology of APA in 1953–1954, and she was active in the affairs of the Society for Research in Child Development, from which she received a distinguished scientific contribution award in 1983. She was president of this society from 1961 to 1963. Nancy Bayley was cited by the American Educational Research Association in 1938 for her 1933 monograph, *Mental Growth During the First Three Years*. She was honored for her extraordinary contributions to the fields of psychology and child development in 1990, in a volume on women in psychology (Lipsitt & Eichorn, 1990).

THE BAYLEY SCALES

The Bayley Mental and Motor Scales were selected for administration to children at 8 months of age in the National Collaborative Perinatal Project. Prior to the adoption of the scales for 8-month use, Bayley supervised the collection of normative data on her scales with one hundred children at each month of age from 1–18 months. A large body of information now exists as a consequence of this normative project. The project

studied children from birth to 7 years of age, born between 1959 and 1966, who survived the first 8 months of life and were returned by their mothers for the 8-month Bayley Scales. Other medical and psychological assessments were also used. Many of the study members between ages 27 and 34 continued to participate in follow-up studies at the University of Pennsylvania and Brown University.

Citing Bayley's paper published ten years before his own celebrated work, James Tanner (1981) said that her seminal paper was the first effort to produce standards for height that took into account an individual's tempo of growth. This was a radical new departure in the approach to standards of growth. He also credited Bayley with the first set of published correlations that related infant size to adult size. She showed life-span predictions to be possible from early size (e.g., girls' adult heights are about double their heights at 2 years of age, whereas boys' heights at 2½ years of age are half their eventual heights) were among Bayley's discoveries from her longitudinal data analyses in the field of human growth. These findings presaged and supplemented her major eventual contribution to the prediction of later mental performance from earlier assessments.

BIBLIOGRAPHY

- BAYLEY, N. (1926). Performance tests for three, four, and five year old children. *Journal of Genetic Psychology*, 33, 435–454.
- BAYLEY, N. (1928). A study of fear by means of the psychogalvanic technique. *Psychological Monographs*, 38, 1–38.
- BAYLEY, N. (1933). *The California First-Year Mental Scale*. Berkeley: University of California Press.
- BAYLEY, N. (1933). Mental growth during the first three years: An experimental study of sixty-one children by repeated tests. *Genetic Psychology Monographs*, 14, 1–92.
- BAYLEY, N. (1951). Some psychologic correlates of somatic androgeny. *Child Development*, 22, 47–60.
- BAYLEY, N. (1955). On the growth of intelligence. *American Psychologist*, 10, 805–818.
- BAYLEY, N. (1958). Value limitations of infant testing. *Children*, 5, 129–133.
- BAYLEY, N. (1968). Behavioral correlates of mental growth: Birth to thirty-six years. *American Psychologist*, 23, 1–17.
- BAYLEY, N. (1993). *The Bayley Scales of Mental and Motor De-*

velopment (revised). San Antonio, TX: The Psychological Corporation.

LIPSITT, L. P., & EICHORN, D. H. (1990). Nancy Bayley. In A. N. O'Connell & N. F. Russo (Eds.), *Women in Psychology: A Biobibliographic Sourcebook*. New York: Greenwood Press.

TANNER, J. M. (1981). *A History of the Study of Human Growth*. Cambridge: Cambridge University Press.

LEWIS P. LIPSITT

BEHAVIORISM Behaviorism, the school of psychology founded by John B. Watson, has had relatively little to say about intelligence. In his founding essay, "Psychology as the Behaviorist Views It," published in 1913, Watson urged psychologists to devote their studies only to phenomena that were directly observable, such as human movements and speech. Subjective phenomena, such as thoughts and images, were, he said, outside the domain of science because they could not be observed by anyone except the person experiencing them, and no one, including that person, could actually measure them with any degree of confidence.

Watson's exhortations were interpreted by some psychologists as license to set aside many, if not all, traditional psychological concepts, particularly if such concepts had any connection whatsoever with the mind, feelings, or the will, the traditional subject matter of psychology since the field's inception in Europe in the 1800s. A great many topics were dubbed useless or uninteresting by various adherents to the behavioristic school, including concepts such as self, ego, traits, intention, purpose, attitude, personality, perception, memory, and thought. Since reasoning, like emotion, takes place in one's head and cannot be observed by others, intelligence, too, was largely ignored, with only a few exceptions.

B. F. Skinner, probably the most influential behaviorist of the twentieth century, proposed a variant of behaviorism he called "radical" behaviorism. The radical form was different from Watson's mainly in allowing private experiences to be studied and analyzed. Curiously, Skinner still considered most traditional psychological concepts, including intelligence, to have

little use, primarily because they distract people, he said, from looking at the role that the environment plays in determining behavior. If a child's poor performance, he argued, is attributed to low intelligence, we might abandon the search for training techniques that might significantly improve the child's performance. The nature–nurture debate over individual differences, he said, was also just a distraction. "The practical question," wrote Skinner (1968), "is not so much whether these differences are genetic or environmental as whether environmental contingencies may be designed to reduce their scope" (p. 241).

Moreover, we make an error of logic, Skinner argued, when we claim to have explained behavior, good or bad, by making reference to high or low intelligence. Intelligence is, he believed, primarily a description or summary of how well people perform on certain tasks. As Isaac Newton warned centuries ago, to use a description or label as an explanation merely creates an illusion of explanation; in no sense does it provide one.

Skinner (1953) acknowledged that tests of intelligence or other traits could indeed be used to make predictions about future performance, but he stressed that in no sense could intelligence be said to be the cause of such performance, but rather that the prediction is "from one effect to another" (p. 199), both performance on an IQ test and subsequent performance being effects of one's genetic endowment and environmental history. The closest Skinner came to suggesting that aspects of intelligence might be worth studying was in his text *The Technology of Teaching* (1968), where he noted that people differ "in speed of learning and forgetting, and as a result in the size of the repertoire that may be acquired and maintained. . . . These," he added, "are presumably the main differences shown by measures of intelligence. Their nature is not clear" (p. 241).

Watson and Skinner have often been said to be extreme "environmentalists"—that is, to believe that all human behavior is learned and that genes play little or no role in individual differences in behavior. In fact, neither they nor most other behaviorists have taken this extreme view. Nevertheless, because many scientists working in the behaviorist tradition have focused their studies on learning processes—in effect, on how behavior is acquired or modified or improved by var-

ious experiences—it is not surprising that they have had little to say about genetic factors. Behaviorists do not actually dismiss genetic factors, but they have historically tended to focus their analyses on the environment. One notable exception to this rule is R. J. Herrnstein, a student of Skinner's, who, in his book *I. Q. in the Meritocracy* (1973), summarized the evidence that genes are important determinants of intelligence.

Arthur Jensen (1984), noted for his strong hereditarian stand on intelligence, has accused behaviorists of claiming that “psychometric tests measure nothing other than the specific bits of knowledge and learned skills reflected in the item content of the tests” (p. 93)—in other words, that IQ tests measure what they test and nothing more. Simplistic thinking of this sort is actually difficult to find in the writings of prominent behaviorists. In any case, Jensen and others have defended the view that intelligence is a general and all-pervasive trait, sometimes represented by the letter *g*. Constructs such as *g* are derived from sophisticated statistical analyses of test scores, especially the technique called factor analysis. Since statistics can usually be interpreted in different ways, and since psychometricians themselves often fail to agree on interpretations, most behaviorists have been wary of this perspective.

Behaviorism is a diverse tradition, with its adherents sometimes having radically different views on the same issue. A few, although very few, individuals who are closely identified with this tradition have written extensively about intelligence. Arthur Staats (e.g., 1963, 1975), proponent of a form of “social behaviorism,” defines intelligence as “specific repertoires—systems of skills—learned according to specified learning principles. . . . The repertoires heavily involve language-cognitive skills, as well as sensorimotor and emotional-motivational basic behavior repertoires. . . . [The repertoires] determine how well the individual will learn, how well the individual will solve problems, and so on” (Staats & Burns, 1981, pp. 241–242; also see Estes, 1974). In other words, intelligence consists of basic skills that have an impact on every aspect of performance, including learning itself and even performance on intelligence tests. Staats has supported this view through a number of experimental studies, mainly with children, that suggest that training in cer-

tain basic skills can significantly improve performance in new situations and, in fact, improve IQ test scores. Staats and Burns (1981) conclude, “Basic behavioral repertoires can be taught to the young child, and . . . this increases specific intelligence test measures” (p. 292).

Also notable is the work on Hans Eysenck of England, who, although highly critical of Skinner's views, is also identified with the behaviorist tradition. Eysenck has written extensively about intelligence and has criticized Skinner for ignoring it. After a debate with Skinner on this topic, Eysenck (1988) remarked, “I had planned to criticise him on the grounds that genetic factors, personality and individual differences generally were excluded from his scheme. He rather took the wind out of my sails by stating explicitly that individual differences, personality, intelligence and their genetic factors were all of very great importance. If this is true, why are they missing from his books, and why does he thunder against those who work in these fields?” (p. 300).

The answer to Eysenck's question is primarily that it is a matter of emphasis. In the early decades of research on learning, most researchers were searching for general laws of learning, and research and theory progressed well without consideration of genetic factors. By the 1970s, with increasingly subtle phenomena under scrutiny in the learning laboratories, exceptions to the laws became commonplace, and genetic factors were needed to account for them. Even Watson, whose early research was on species-specific behavior in animals, rejected a simplistic environmentalist view in some of his writings, and Skinner wrote several essays about genetic issues in his later years.

Outside the boundaries of the Watson-Skinner tradition in psychology, the pragmatist philosopher George Herbert Mead and the “interbehaviorist” J. R. Kantor offered analyses of intelligence in behavioral terms. Of special note is the work of personality theorist Walter Mischel (e.g., 1981), who has shown that behavior often attributed to traits is affected in orderly ways by the situations people face: “If you want to predict what somebody's going to do in a particular situation now . . . , probably the best estimate will be made from the closest, single approximation of behavior in that situation” (1981, p. 92). In other words, as

Skinner argued, the environment plays an important role in determining what people do.

(See also: LEARNING AND INTELLIGENCE; THURSTONE, L. L.)

BIBLIOGRAPHY

- ESTES, W. K. (1974). Learning theory and intelligence. *American Psychologist*, 29, 740–749.
- EYSENCK, H. J. (1988). Skinner, Skinnerism, and the Skinnerian in psychology. *Counselling Psychology Quarterly*, 1, 299–301.
- HERRNSTEIN, R. J. (1973). *I. Q. in the meritocracy*. Boston: Little, Brown.
- JENSEN, A. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures*, 7, 93–118.
- MISCHEL, W. (1981). Current issues and challenges in personality. *G. Stanley Hall Lecture Series*, 1, 81–99.
- SKINNER, B. F. (1953). *Science and human behavior*. New York: Free Press.
- SKINNER, B. F. (1968). *The technology of teaching*. New York: Appleton.
- STAATS, A. W. (WITH CONTRIBUTIONS BY C. K. STAATS). (1963). *Complex human behavior*. New York: Holt, Rinehart & Winston.
- STAATS, A. W. (1975). *Social behaviorism*. Homewood, IL: Dorsey Press.
- STAATS, A. W., & BURNS, G. L. (1981). Intelligence and child development: What intelligence is and how it is learned and functions. *Genetic Psychology Monographs*, 104, 237–301.
- WATSON, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, 20; 158–177.

ROBERT EPSTEIN

BENTON, ARTHUR L. (1909–) Arthur Lester Benton was born in New York City, on October 16, 1909. In 1931, he received his B.A. from Oberlin College, and he earned his master's there in 1933. He received his doctorate in psychology from Columbia University in 1935.

Arthur Benton's work became one of the cornerstones of an international interdisciplinary effort that founded and developed the field of human neuropsy-

chology as a science. In turn, this helped establish clinical neuropsychology as a recognized specialty within professional psychology. Benton gained international renown for his contributions as a scientist, practitioner, scholar, teacher, organizer, and leader.

Broadly speaking, neuropsychology is the study of brain–behavior relationships. The particular interest of neuropsychology is the complex and distinguishing behavior of humans that includes cognition, conation (desire and volition), and affect. The beginnings of neuropsychology were ushered in by the cautious conclusion of Paul Broca (1865) in the 1850s and 1860s that an acquired disorder of expressive language could be localized to a lesion in a specific part of the brain—the left side of the cerebrum. This helped establish the doctrine of “hemispheric cerebral dominance,” meaning that the two halves of the cerebrum were specialized for different cognitive abilities. At about the same time, the doctrine of “localization of function” became important—that vision, motor movement, tactile sensation, and so on, were controlled by different parts of the cerebrum. Combined, these two doctrines implied that at least some mental phenomena were the product of discrete, almost tangible, structural and physiological mechanisms in the brain. Since this inquiry involved assessing patients with brain disease and localizing their lesions by autopsy, it was mainly conducted in neurological centers. There followed many reports of mental phenomena attributed to disease in specific parts of the cerebrum; however, these demonstrations often were not standardized or adequately developed, and they were seldom tested on control subjects. Therefore, it was difficult to distinguish valid from invalid findings and to resolve contradictory conclusions.

Benton's formative interest in this field was shaped by two professors at Oberlin College during his undergraduate and master's studies. Later, as a doctoral student in psychology at Columbia University, his interests were specialized toward the psychological testing of patients with various types of brain disease. In the late 1930s, as a practicing pediatric psychologist, his appreciation of the types of specific cognitive deficits that could result from brain disease increased, and he identified a need for a reliable nonverbal test of perceptual memory. Also, he began experimenting

with a test of finger localization; this study continued while he served in the U.S. Navy during World War II. His work had introduced him to the variety of visual and other perceptual disturbances that resulted from brain disease. In 1945, Benton published the first version of his now internationally famous Visual Retention Test.

Benton's overarching interest in the mind, the brain, and measurement can be seen in his earliest writings, starting with publications in the 1930s on the development and application of psychological tests for clinical use. In 1940, he presented a substantial critical review of the mental development of premature children. In 1941, he demonstrated how psychological tests could be used to identify a head-injured patient's presumed "neurotic" symptoms as manifestations of brain trauma. After World War II, Benton became an associate professor at the University of Louisville School of Medicine. He published an extensive case report on the psychological effects of bilateral frontal lobe disease in 1947. In 1948, he was brought to the University of Iowa to head the new doctoral program in clinical psychology. In a little more than a year, he developed associations within the medical school and established a research laboratory and neuropsychology service in the Department of Neurology at the University of Iowa Hospitals and Clinics. The identification and localization of cerebral disease through the evaluation of mental functions figured prominently in the diagnostic roster of neurological medicine. Consequently, neurology and neurosurgery welcomed Benton's expertise and collaboration.

It was the melding of the scientific rigor of experimental psychology with the careful development of specific psychological tests to investigate the behavioral consequences of brain disease that helped launch the contemporary field of human neuropsychology—and Benton's own international preeminence. Benton's research program stressed the importance of understanding the historical roots of neuropsychological issues and of maintaining a sense of historical continuity in conducting and interpreting empirical investigations.

Despite the work of Benton and others, older traditions persisted in the medical literature. In 1961, Benton issued in the *Journal of Neurology, Neurosurgery and Psychiatry* a startling denunciation of the then

widely accepted "Gerstmann syndrome," calling it a fiction. The "syndrome" consisted of four behavioral signs of brain disease (right–left disorientation, acalculia, agraphia, and finger agnosia), which Gerstmann believed to be the manifestations of a single problem (a disturbance in body schema). The presence of the syndrome was said to identify a lesion in a particular area of the brain (the angular gyrus in the left parietal lobe). Benton demonstrated through research and review of the literature that this syndrome was "an artifact of defective and biased observation." More important, he warned that the uncritical acceptance of scientifically unfounded conclusions "carries the hazard of retarding advances in the understanding of the organization of abilities and disabilities in patients with cerebral disease" (p. 181). This led to the recognition that a new standard of evidence, one that was more psychologically sophisticated and scientifically rigorous, was required for the proper study of brain–behavior relations.

The history of neuropsychology and behavioral topics in neurology remained a continuing interest of Arthur Benton. Some of the major areas of investigative research in his program included the method of double tactile stimulation, motor imperistence, aphasic disorders, and asymmetries in hemispheric functions as reflected in diverse visual, constructional, tactile, and auditory types of performance. Benton charted the development of right–left orientation and finger recognition in normal children and in children with cognitive disorders. This work demonstrated how different cognitive abilities were involved in different types of performance and how the prominence of each component varied with and depended upon the child's stage of mental development. Findings such as these, along with reviews of the behavioral effects of various types of cerebral disease in children, helped identify fundamental and specialized considerations for research and clinical practice in child neuropsychology. Reaction time in patients with cerebral disease, including the effects of motivational factors and task complexity, was another major topic associated with the Benton Laboratory of Neuropsychology. A more detailed account can be found in a paper by Hamsher (1985).

Benton was a founding organizer of the International Neuropsychological Society. He held appoint-

ments as a visiting scientist and scholar at universities around the world (Eron, 1992). He has been recognized for his distinguished contributions to both the science and profession of neuropsychology by the American Psychological Association, the American Psychological Foundation, the American Board of Professional Psychology, the International Neuropsychological Society, the Orton Dyslexia Society, the American Board of Clinical Neuropsychology, the National Academy of Neuropsychology, and others.

Beginning in 1948 and continuing for over four decades, for psychology graduate students aspiring to specialize in neuropsychology, in his laboratory and in the University of Iowa Hospitals and Clinics, Benton provided a tailored program of study and experiences. Visiting scientists frequented his laboratory and enriched the process. This evolved into a curriculum for the making of a new professional—the clinical neuropsychologist. Benton was active in the development of the definition of the field when it was an emerging specialty, in the development of guidelines for education and training, and in establishing a board for certifying competent practitioners.

Tests devised by Benton to answer specific questions or meet particular needs in research often proved to hold clinical significance and utility in their own right; therefore, they were revised and standardized for clinical application. For example, prosopagnosia is a rare but fascinating neurobehavioral syndrome in which the affected patient loses the ability to identify familiar people by their faces despite having adequate vision. To test the hypothesis that this syndrome might be the severe manifestation of a more common deficit that could be found in other patients with brain disease, a test for the recognition of unfamiliar faces was constructed. Evidence from Benton's and other research laboratories failed to support an important connection between this type of performance and prosopagnosia. Nevertheless, in the process, this test was found to be a sensitive indicator of certain types of cognitive dysfunction in patients with focal brain lesions.

Benton's other contributions to neuropsychological assessment included the Neurosensory Center Comprehensive Examination for Aphasia (Spreeen & Benton, 1969), the Multilingual Aphasia Examination (Benton & Hamsher, 1989), and an array of specialized tests of

orientation, learning, perception (visual, spatial, tactile, and auditory), and motor imperistence (Benton et al., 1983). To provide in a single source an overview of Benton's contributions to neuropsychology, Costa and Spreeen (1985) reprinted a selected collection of his historical, theoretical, and empirical papers.

BIBLIOGRAPHY

- BENTON, A. L. (1961). The fiction of the "Gerstmann Syndrome." *Journal of Neurology, Neurosurgery and Psychiatry*, 24, 176–181.
- BENTON, A. L., & HAMSHER, K. DES. (1989). *Multilingual Aphasia Examination* (2nd ed.). Iowa City: AJA Associates.
- BENTON, A. L., HAMSHER, K. DES., VARNEY, N. R., & SPREEEN, O. (1983). *Contributions to neuropsychological assessment*. New York: Oxford University Press.
- BROCA, P. (1865). Du siège de la faculté du langage articulé [The seat of the faculty of articulate language]. *Bulletin de la Société d'Anthropologie*, 6, 337–393.
- COSTA, L., & SPREEEN, O. (Eds.). (1985). *Studies in neuropsychology: Selected papers of Arthur Benton*. New York: Oxford University Press.
- ERON, L. D. (1992). APF Gold Medal Awards and Distinguished Teaching of Psychology Award. *American Psychologist*, 47, 851–853.
- HAMSHER, K. DES. (1985). The Iowa group. *International Journal of Neuroscience*, 25, 295–305.
- SPREEEN, O., & BENTON, A. L. (1969). *Neurosensory Center Comprehensive Examination for Aphasia*. Victoria: University of Victoria.

KERRY DES. HAMSHER

BIAS IN TESTING Bias as a common statistical term refers to systematic, predictable, directional error, as opposed to the random error present in all psychological and educational tests of less than perfect reliability (i.e., all tests). In more common usage, bias may refer to warped or prejudicial judgment or a leaning toward such biased judgments based upon attitudes or interests that may be unfounded. The two usages of the term *bias* are easily confused; the former often is unknown to the lay person, and some psychologists and educators simply fail to make the distinction between the two usages. In the present entry, *bias* is used according to the statistical definition given

above, as it must be if it is to be used to investigate potential bias in tests of intelligence and other psychological instruments of measurement.

A test may be biased along many dimensions. Most often, such variables as gender, race, ethnicity, and socioeconomic status are cited as dimensions on which tests, especially of intelligence, may be very biased. This has led to the development of the "cultural test bias" hypothesis, the most prominent and emotionally laden criticism relating to bias in testing (Reynolds & Kaiser, 1990).

The cultural test bias hypothesis is the contention that differences in the mental test scores of members of different racial and ethnic groups are the result of inherent flaws in the tests themselves. These flaws bias, or cause systematic error, in a manner that causes ethnic minorities to earn low scores. It is therefore believed that the mean differences between groups tested on such things as intelligence quotient (IQ) are artifacts of the test and do not reflect any real differences in mental abilities or skills.

Mean differences in mental test scores between races are some of the most frequently published findings in psychological research on individual differences. One of the primary explanations of differences in scores is that the people tested have been reared in very different environments, with lower-scoring groups having been relatively deprived of the quantity and quality of stimulation received in the formative years by the higher-scoring groups. Another explanation is that the measured ability of the lower-scoring groups reflects a difference in their genetic potential for intellectual performance. However, most contemporary authors take the position that the lower scores earned by some racial and ethnic groups are the result of some still-undefined interplay of environment and genetics.

Cultural bias in testing has existed as a potential explanation at least since it was raised by Sir Cyril BURT (1921) and by R. Pintner and R. Keller (1922), with occasional papers on the issue appearing over the years. It was not widely accepted as a serious hypothesis until the late 1960s, when the Association of Black Psychologists (ABP) called for a moratorium on the psychological testing of minorities and disadvantaged students, particularly for the purpose of placing such children in special education programs. In 1969 the ABP issued an official policy statement encouraging

parents of black children to refuse to allow their children or themselves to be evaluated on any achievement, intelligence, aptitude, or performance test.

The primary objections to the testing of minorities because of race or cultural bias in the tests have been classified by C. R. Reynolds (1982a; 1987) into six categories as follows:

1. *Inappropriate content.* Black and other minority children have not been exposed during their development to the material dealt with in the test questions or other stimulus materials. Rather, the item content of such tests is geared toward white middle-class homes and values.

2. *Inappropriate standardization samples.* During the development of tests of mental abilities, ethnic minorities are underrepresented in the collection of normative reference-group data. It was not unusual several decades ago for all standardization samples of major tests to be white only.

3. *Examiner and language bias.* Since most psychologists who administer such tests are white and speak only standard English, they intimidate blacks and other minorities. They are also unable to communicate accurately with minority children. Lower test scores for minorities, then, are said to reflect this intimidation and difficulty in the communication process, not lower ability levels.

4. *Inequitable social consequences.* As a result of bias in educational and psychological tests, minority-group members, who are already at a disadvantage in the educational (and later vocational) markets because of past discrimination, are disproportionately directed onto dead-end educational tracks and so are thought unable to learn. "Labeling effects" also fall into this category.

5. *Measurement of different constructs.* Related to the question of inappropriate content, this objection asserts that today's tests are still measuring significantly different attributes when used with children from other than the white middle-class culture.

6. *Differential predictive validity.* While tests may accurately predict a variety of outcomes (such as numbers of years of schooling completed and high school and college grade point averages) for white middle-class children, they fail to predict at an acceptable level any real-life outcomes for minority-group members that are relevant for those groups. A corollary to this

objection is a variety of competing positions regarding the selection of an appropriate, common criterion against which to validate tests across cultural groupings. Scholastic or academic attainment levels also are considered by a variety of black psychologists as biased criteria.

The actions by the ABP had several positive effects. Prior to the association's call for a moratorium on the testing of minorities, little actual research existed in the area. Subsequently, much research was prompted by the ABP position as it brought the race bias hypothesis to the forefront of explanations of race differences in intelligence. Also in response to this call for a moratorium, the American Psychological Association's Board of Scientific Affairs appointed a committee to study the use of tests with disadvantaged students. The committee, headed by T. Anne Cleary, gave its official report in the form of an article in the *American Psychologist* (Cleary, Humphreys, Kendrick, & Wesman, 1975). After an extensive review of many potential sources of bias, the authors concluded that "in general, more and better data are recommended as one of the most desirable elements in a program concerned with better and fairer use of tests" (p. 40).

Research on race bias in testing was, and continues to be, of major importance to psychologists as well as to society at large. The cultural test bias hypothesis is a major scientific question facing psychology today (Reynolds, 1981). If this hypothesis is ultimately accepted as correct, then the past century or so of research into the psychology of individual differences (or differential psychology, the basic psychological science underlying all fields of applied psychology) will have been confounded and will have to be reexamined. Nevertheless, race bias in testing has been, and continues to be, tested both in the judicial courts and in the scholarly court of open inquiry. Three major federal court cases have decided the issue of cultural bias in intelligence testing. The first and most famous of these cases, known as the *Larry P.* case (1979; Reynolds & Mann, 1987), resulted in a decision that standardized intelligence tests were in fact biased against black children and could not be used for placement in special education. However, the *Larry P.* decision was subsequently overturned and two additional, highly similar cases (*PASE v. Hannon*, 1980; *Marshall v. Georgia*, 1984) yielded opinions indicating that intelligence tests are

not biased against racial minorities. These three cases are discussed at length in C. R. Reynolds and L. Mann (1987) and D. Bersoff and P. Hofer (1990).

Only since the mid-1970s has considerable research been published regarding race bias in testing. For the most part, this research has failed to support the test bias hypothesis, revealing instead that (1) well-constructed, well-standardized educational and psychological tests predict future performance in an essentially equivalent manner across race for American-born ethnic minorities; (2) the internal psychometric structure of the tests is essentially not biased in favor of one race over another; and (3) the content of the items contained in these tests is about equally appropriate for all these groups. Some have challenged certain assumptions underlying these investigations, so we do not now have definitive answers. On a related issue, reviews of the literature on bias in employment testing by J. E. Hunter, F. L. Schmidt, and R. Hunter (1979) and of bias in personality assessment by M. P. Moran (1990) have also drawn quite similar conclusions.

The controversies of the past have prompted the major test-publishing houses to investigate in depth the potential ethnic, racial, and gender bias of tests prior to publication, namely, during the test-development phase, when necessary changes still can be made. Using a variety of bias-detection statistical methods, most publishers assay not only individual items but also entire scales. Some prominent examples of tests by various publishers that evaluated bias on an a priori basis include *BASC: Behavior Assessment System for Children* (Reynolds & Kamphaus, 1992), *Test of Memory and Learning* (Reynolds & Bigler, 1993), and *Wechsler Intelligence Scale for Children-III* (Wechsler, 1991). The manual for each of these tests contains details of the various analyses done to evaluate potential bias during the developmental phase of the measure. Ultimately, such efforts will improve the practices of minority testing, because test publishers and consumers have been sensitized to the issues of bias to a degree unknown prior to the mid-1970s.

The above comments notwithstanding, race bias in testing is one of the most controversial and emotional issues in psychology. It will not be resolved entirely on the basis of published research findings, because in the past many tests have unquestionably been abused

when given to minority groups. Much of the current controversy still centers on the placement of larger numbers of minority children in special education programs. Thus, special care must be taken by test developers and publishers to ensure that the misuses and abuses of the past are thwarted by "intelligent testing" (Kaufman, 1979).

A general review of race bias in testing can be found in A. R. JENSEN (1980). Specialty reviews have been published by Hunter, Schmidt, and Hunter (1979) on race bias in employment testing and by Reynolds (1982a) on bias in the testing of children. A book-length debate covering a number of relevant issues is found in Reynolds and Brown (1984). Methodology for investigating most aspects of cultural bias in testing relevant to special education is reviewed in Jensen (1980) and in Reynolds (1982b).

(See also: ETHNICITY, RACE, AND MEASURED INTELLIGENCE.)

BIBLIOGRAPHY

- BERSOFF, D., & HOFER, P. (1990). The legal regulation of school psychology. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology*. New York: Wiley.
- BURT, C. (1921). *Mental and scholastic tests*. London: P.S. Kiy.
- CLEARY, T. A., HUMPHREYS, L. G., KENDRICK, S. A., & WESMAN, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist*, 30, 15–41.
- HUNTER, J. E., SCHMIDT, F. L., & HUNTER, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, 86, 721–735.
- JENSEN, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- KAUFMAN, A. S. (1979). *Intelligent testing with the WISC-R*. Interscience. New York: Wiley.
- LARRY P. V. W. RILES, 495 F. Supp. 926 (N.D. Cal. 1979).
- MARSHALL V. GEORGIA, CV 483–233 (U.S.D. S.D. Geo., 1984).
- MORAN, M. P. (1990). The problem of cultural bias in personality assessment. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Vol. 2. Personality, behavior, and context*. New York: Guilford Press.
- PASE V. HANNON, 506 F. 2d 831 (N.D. Ill. 1980).
- PINTNER, R., & KELLER, R. (1922). Intelligence tests of foreign children. *Journal of Educational Psychology*, 13, 214–222.
- REYNOLDS, C. R. (1981). In support of bias in mental testing and scientific inquiry. *Behavioral & Brain Sciences*, 3, 352.
- REYNOLDS, C. R. (1982a). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology*. New York: Wiley.
- REYNOLDS, C. R. (1982b). Methods for detecting construct and predictive bias. In R. A. Berk (Ed.), *Handbook of methods for detecting test bias*. Baltimore: Johns Hopkins University Press.
- REYNOLDS, C. R. (1987). Cultural bias in testing. In C. R. Reynolds & L. Mann (Eds.), *Encyclopedia of special education*. New York: Wiley.
- REYNOLDS, C. R., & BIGLER, E. D. (1993). *Test of memory and learning*. Austin, TX: PRO-ED.
- REYNOLDS, C. R., & BROWN, R. T. (1984). *Perspectives on bias in mental testing*. New York: Wiley.
- REYNOLDS, C. R., & KAISER, S. M. (1990). Test bias in psychological assessment. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (2nd ed.). New York: Wiley.
- REYNOLDS, C. R., & KAMPHAUS, R. W. (1992). *BASC: Behavior assessment system for children*. Circle Pines, MN: American Guidance Service.
- REYNOLDS, C. R., & MANN, L. (EDS.) (1987). *Encyclopedia of special education*. New York: Wiley.
- WECHSLER, D. (1991). *Wechsler intelligence scale for children-III*. San Antonio, TX: The Psychological Corporation.

Cecil R. Reynolds

BILINGUALISM The present state of understanding about bilingualism can be traced to three principal areas of research. The first is the question of the relationship of bilingualism to intelligence, a literature that dates back to the origins of psychometrics and extends to present-day cognitive approaches. The second is the study of second-language acquisition as defined by a combination of questions from language teaching (e.g., how much attention to give to the native language of the learners) and inspiration from rationalist approaches to language that followed the collapse of behaviorist accounts. The third is the level-of-analysis issue of whether to conceptualize bilin-

gualism as a psycholinguistic or a sociolinguistic phenomenon.

BILINGUALISM AND INTELLIGENCE

The question of whether bilingualism resulted in a "language handicap" on standardized tests of intelligence can be found in the literature from the early twentieth century. The key tension in this literature was whether bilingualism caused the poor performance of immigrant children. Those who favored "nature" explanations claimed that the language handicap itself was the result of hereditary factors (Goodenough, 1926). In contrast, advocates of the "nurture" explanation saw the experience of bilingualism as causing mental hardship and linguistic confusion (Smith, 1931).

More recent researchers found positive effects of bilingualism and pointed to methodological and sociological problems associated with the early research. On the methodological side, E. Peal and W. E. Lambert (1962) noted that the selection criteria for the early research did not assess for bilingualism and that a fairer assessment of bilinguals can be made by selecting for "balanced bilinguals," those with equivalent proficiencies in the two languages. On the sociological side, it has been noted that the early research focused almost exclusively on immigrant and lower-status bilinguals, excluding populations for whom bilingualism resulted in enhanced social status (Fishman, 1977).

Research has also expanded the dependent variable, moving away from narrow conceptions of intelligence to a wide array of measures, such as specific thinking skills, creativity, social cognition, and metalinguistic awareness (see Reynolds, 1991). A clear generalization is that when subjects are selected on the basis of being balanced bilinguals, they perform at a level at least equivalent to monolingual controls, and in many cases, the results show a positive effect of bilingualism, although the effect sizes are small to moderate. These effects are demonstrable even in low-status bilingual children as long as their degree of bilingualism is controlled (Diaz, 1985).

The major challenge to this field of knowledge is more theoretical than empirical. The emphasis thus far has been on demonstrating the effect, with a corresponding lack of attention to providing an explanation

of the effect. Aside from the area of metalinguistic awareness in which there is a hypothesized link to automatic versus controlled processing (Bialystok, 1988) and to Lev VYGOTSKY's theory of word-object separation (Janco-Worrall, 1972), there has been very little activity in the field. Now that fears about the negative consequences of bilingualism have been put to rest, the field would be well served by strong linkages to more general theories of language and cognition that might explain the positive effects obtained.

SECOND-LANGUAGE ACQUISITION

The main questions surrounding second-language acquisition are the role of the native language and determinants of individual differences in successful acquisition. Up until the 1960s, second-language acquisition and teaching were defined by the contrastive analysis of grammatical structures of the native (first) and target (second) languages, in which potential sources of positive and negative transfer were identified (Lado, 1964). This view of second-language acquisition was rooted in empiricist accounts of language and learning and became discredited in the face of the popularity of the rationalist views of language acquisition that followed Noam Chomsky (1957).

Studies of second-language acquisition followed the path of studies of child language acquisition, in which the errors produced by the learner were considered an important window into the developmental process. Studies of second-language learner errors conducted in the 1960s and 1970s supported the general move away from contrastive analysis, in that a remarkably small proportion of errors observed in the learner speech could be traced to the native language (Larsen-Freeman & Long, 1990). The majority of errors were similar to those found in child language learners, including the simplification of structures (e.g., omission of grammatical inflections) and overgeneralization of rules (e.g., past tense *-ed* added to irregular verbs as in *eated*, instead of *ate*).

Although it is clear that source-language errors are rare in second-language learners, this does not mean that the source language is unimportant. First, there are persistent reports of difficulty in specific areas of grammar that are related to the source language, such as the English article system for speakers of many

Asian languages. Second, within the framework of linguistics called "universal grammar," which is distinctly rationalist in orientation, predictions are made that certain abstract linguistic parameters that distinguish between groups of languages are "set" in the process of first-language acquisition. This setting may have consequences for second-language acquisition, depending on whether the parameter in the target language is the same as, or different from, the native language. Although the empirical tests of this theory are still being worked out (White, 1989), there is a strong theoretical attempt to revive the role of the source language within a rationalist framework. And third, as trainers of foreign languages know, different languages take vastly different amounts of training to master. The Foreign Service Institute, for example, offers intensive language courses of very different lengths, depending on the language, ranging from twenty weeks for French, German, Italian, and Spanish to forty-four weeks for Arabic, Turkish, and Urdu (Odlin, 1989). These differences have much to do with each language's degree of similarity to English. These are clear influences of the native language that are obscured when one only pays attention to the process of second-language acquisition.

With respect to the question of individual differences, language aptitude is clearly important (Carroll, 1981), as are attitude and motivation (Gardner, 1985). The role of the social-psychological variables is especially evident in the case of language learning that occurs in settings where the second language is not a prominent feature of the sociolinguistic landscape, such as in the learning of French in the English-speaking parts of Canada or the learning of most foreign languages in American classrooms. Some have speculated about the relevance of other features, such as personality and cognitive style, but these effects are far from established.

Another important source of individual variation is the age at which second-language learning begins. In a review of the literature, Long (1990) found that there is sufficient evidence to conclude that there are maturational constraints on second-language acquisition, that is, the younger the learner, the better, particularly in the areas of phonology, morphology, and syntax. It is difficult to interpret this age effect as due

to biological constraints, such as a critical period, as hypothesized by E. H. Lenneberg (1967). First, age effects are generally linear in nature and appear both before and after the supposed end of the critical period at puberty. Second, considerable within-age variation and many instances of highly successful adult second-language acquisition have been documented. Third, instances of qualitative differences in the grammatical development of child and adult second-language acquisition have not been documented.

PSYCHOLINGUISTIC AND SOCIOLINGUISTIC PERSPECTIVES

A psycholinguistic definition of bilingualism results from examining the question of an individual's relative proficiency in the two languages. In contrast, a sociolinguistic definition results from studying the question of the speech communities to which the bilingual individual belongs.

The psycholinguistic perspective has resulted in the classification of individuals into compound versus coordinate bilingualism (Weinreich, 1953), balanced versus unbalanced bilingualism (Peal and Lambert, 1962), and early versus late bilingualism (Genesee et al., 1978). Although each of these classifications results from different sociolinguistic experiences, each is thought to result in distinct psychological organizations that would have measurable consequences in psycholinguistic behavior. The compound versus coordinate distinction has been subjected to considerable empirical scrutiny. According to this distinction, the lexicon is organized either on the basis of a single concept associated with the corresponding words in the two languages or on the basis of separate concepts for each language. None of the evidence produced thus far validates the distinction. The null hypothesis holds that if one is bilingual, it does not matter how one got there.

The sociolinguistic perspective has produced distinctions along the lines of the social status of the languages involved. These include elite versus folk bilingualism, referring to whether bilingualism is a marker of elite or plebeian social status (Fishman, 1977); additive versus subtractive bilingualism, referring to whether the second language enriches or

threatens the native language (Lambert, 1975); and elective versus circumstantial bilingualism, referring to whether bilingualism is a consequence of individual choice or an accompaniment of a social reality, such as immigration or annexation (Valdes, 1992). Such distinctions help account for whether bilingualism is valued and maintained or allowed to shift into monolingualism. They also explain the language policies adopted by the government and educational systems toward bilingualism.

Psycholinguistic and sociolinguistic dimensions of bilingualism are in principle separable from one another. For example, in the United States many individuals with proficiency in English and another language spend most of their lives as practicing English monolinguals (Veltman, 1983). In a sense, these individuals are psycholinguistically bilingual but sociolinguistically monolingual. This distinction is especially useful in understanding what is happening to the native language of minority communities in the United States and other immigrant countries. There is strong evidence of a rapid shift toward a preference for English among immigrants, as indicated by census information. This shift is not a psycholinguistic phenomenon, the result of individuals losing their bilingual proficiency in the course of their lifetime; rather, it is a sociolinguistic phenomenon, the result of the low-status immigrant language falling into disuse and then failing to be transmitted from one generation to the next (Hakuta & D'Andrea, 1992). The psycholinguistic perspective, then, tells us how languages are learned, but it is the sociolinguistic perspective that tells us how a language is lost by a community.

CONCLUSION

The concept of intelligence has undergone major changes since the question of the language handicap was first raised. Likewise, the concept of bilingualism has become far more complex than the simple measurement of vocabulary or grammar in the two languages. The issue is no longer whether the theories need to bring social factors into account but rather the manner in which linguistic and cognitive theories interact with social theories. At this point, there is little evidence to suggest that the linguistic and cognitive

aspects of bilingualism are affected in any qualitative ways by the social factors involved.

(See also: LANGUAGE AND INTELLIGENCE.)

BIBLIOGRAPHY

- BIALYSTOK, E. (1988). Levels of bilingualism and levels of linguistic awareness. *Developmental Psychology* 24, 560–567.
- CARROLL, J. (1981). Twenty-five years of research on foreign-language aptitude. In K. C. Diller (Ed.), *Individual differences and universals in language learning aptitude*. Rowley, MA: Newbury House.
- CHOMSKY, N. (1957). *Syntactic structures*. The Hague: Mouton.
- DIAZ, R. M. (1985). Bilingual cognitive development: Addressing three gaps in current research. *Child Development*, 56, 1376–1388.
- FISHMAN, J. (1977). The social science perspective. In *Bilingual education: Current perspectives: Social science* (pp. 1–49). Arlington, VA: Center for Applied Linguistics.
- GARDNER, R. C. (1985). *Social psychology and second-language learning*. London: Arnold.
- GENESE, F.; HAMERS, J.; LAMBERT, W. E.; MONONEN, L.; SEITZ, M.; & STARCK, R. (1978). Language processing in bilinguals. *Brain and Language*, 5, 1–12.
- GOODENOUGH, F. (1926). Racial differences in the intelligence of school children. *Journal of Experimental Psychology*, 9, 388–397.
- HAKUTA, K. (1983). English language acquisition by speakers of Asian languages. In M. Chu-Chang (Ed.), *Asian and Pacific-American perspectives in bilingual education* (pp. 31–55). New York: Teachers College Press.
- HAKUTA, K. (1986). *Mirror of language: The debate on bilingualism*. New York: Basic Books.
- HAKUTA, K., & D'ANDREA, D. (1992). Some properties of bilingual maintenance and loss in Mexican background high school students. *Applied Linguistics*, 13, 72–99.
- IANCO-WORRALL, A. (1972). Bilingualism and cognitive development. *Child Development*, 43, 1390–1400.
- LADO, R. (1964). *Language teaching: A scientific approach*. New York: McGraw-Hill.
- LAMBERT, W. E. (1975). Culture and language as factors in learning and education. In A. Wolfgang (Ed.), *Education*

- of immigrant students* (pp. 55–83). Toronto: Ontario Institute for Studies in Education.
- LARSEN-FREEMAN, D., & LONG, M. (1990). *An introduction to second-language acquisition research*. London: Longman.
- LENNEBERG, E. H. (1967). *Biological foundations of language*. New York: Wiley.
- LONG, M. (1990). Maturation constraints on language development. *Studies in Second-Language Acquisition*, 12, 251–285.
- ODLIN, T. (1989). *Language transfer: Cross-linguistic influence in language learning*. Cambridge: Cambridge University Press.
- PEAL E., & LAMBERT, W. E. (1962). The relation of bilingualism to intelligence. *Psychological Monographs*, 76 (27, Whole No. 546).
- REYNOLDS, A. G. (1991). The cognitive consequences of bilingualism. In A. G. Reynolds (Ed.), *Bilingualism, multiculturalism, and second-language learning: The McGill conference in honour of Wallace E. Lambert*. Hillsdale, NJ: Erlbaum.
- SMITH, M. (1931). Some light on the problem of bilingualism as found from a study of the progress in mastery of English among pre-school children of non-American ancestry in Hawaii. *Genetic Psychology Monographs*, 21, 119–284.
- VALDES, G. (1992). Bilingual minorities and language issues in writing. *Written Communication*, 9, 85–136.
- VELTMAN, C. (1983). *Language shift in the United States*. New York: Mouton.
- WEINREICH, U. (1953). *Languages in contact: Findings and problems*. New York: Linguistic Circle of New York.
- WHITE, L. (1989). *Universal grammar and second-language acquisition*. Amsterdam: John Benjamins.

KENJI HAKUTA

BINET, ALFRED (1857–1911) Alfred Binet was born in Nice, France, in 1857, and considered following his father and grandfather into the profession of medicine. He decided otherwise when his physician-father had him touch a cadaver to prepare him for his profession. Instead, he studied law and in 1878, at age 21, took his license in it. He never practiced law (or any other profession, owing to his independent wealth) but began to study psychology on his own at the Bibliothèque Nationale (the National Library). At age 23, he published his first paper on the “fusion” that occurred in an individual’s perception upon being

simultaneously stimulated (i.e., pricked by a pin) at two different points of the body separated by various distances. Between 1882 and 1883 Binet gravitated to Professor C. H. Fere and Professor Jean Charcot at the Salpêtrière and, during the next seven years, published numerous papers with Fere. The papers described the powerful influence magnets allegedly had on the behavior of hypnotized patients, including transferring perceptions of movements from one side of the body to the other in consort with the position of the magnet. The papers also described the capacity of magnets to initiate and influence illusions, perceptions, and hallucinations—and even to reverse emotions (such as hate into love, joy into despair). This research was severely criticized by Professor Joseph R. L. Delboeuf, who visited and found Binet and Fere, two inexperienced investigators, openly announcing in front of their hypnotized patient exactly which results they expected to achieve under hypnosis. Delboeuf returned to his own laboratory in Liege, repeated the experiments with the proper controls and, in 1886, published his negative findings—showing that Binet and Fere had been duped.

From 1887 to 1891, Binet went to study and carry out research in the laboratory of his embryologist father-in-law, Professor Edward G. Balbiani. Binet was awarded a doctorate in natural science from the Sorbonne in 1894, based on his search for anatomical-physiologic correlates of behavioral responses in insects. Earlier, in 1891, continuing his own study of psychology, Binet had gone to work at the Sorbonne for Professor Henri E. Beaunis who, in 1889, had founded that university’s Laboratory of Physiological Psychology. Before beginning to work with Beaunis, and three years before completing his doctoral studies in natural science, Binet had begun the study of child psychology through the systematic investigation of the developmental processes observed in his own two daughters—Madeleine, born December 1885, and Alice, born July 1887. In the first paper describing that research (published in 1890; reprinted in Pollack & Brenner, 1969), Binet reported that children could *not* be distinguished from adults in the nature of their responses to *simple* measures of reaction time and other physiological indices. In that same 1890 paper on child versus adult functioning, Binet flirted with what later

would be his discovery of the concept of a person's *mental age* when he suggested:

If one could succeed in measuring intelligence, that is, reasoning, judgment, memory, the power of abstraction—and this does not appear to me absolutely impossible—the number expressing the mean intellectual development of an adult would represent a relationship quite different from the number expressing the intellectual development of a child (Binet, 1890, p. 74; Pollack & Brenner, 1969, p. 85).

Thus, while studying complex (as opposed to simple) mental operations in his two young daughters and while also completing his dissertation, Binet had the first glimpse of what he later would call *mental age* and its almost full development before the end of adolescence.

Despite the flourishing German and American emphasis on simple sensory processes, the next decade would see Binet progressively refine his feebly emerging conception of intelligence as a characteristic of global human performance—a unitary characteristic that is present in young children and can be assessed, even in them, by questions requiring answers exhibiting complex acts of judgment or reasoning. As a forerunner of Jean PIAGET, Binet began to appreciate the stages of reasoning through which the average child develops. In the 1890 publications (Wolf, 1966; see the full translations in Pollack & Brenner, 1969) are examples of the types of judgment and reasoning items that would appear in the 1905 scale (e.g., “What does it mean to be afraid?”), as well as Binet's appreciation of the complex phenomena (1) that such items could tap and (2) whose underlying psychological and theoretical properties Piaget would begin to unravel many decades later (Pollack, 1971).

By 1892, Binet became codirector and then formally succeeded Beaunis in 1895 as director of the Laboratory of Physiological Psychology. Nevertheless, Binet was not yet able to translate his observations into a viable approach for assessing intellectual ability. In 1896, Binet and V. Henri published a paper describing a program of projected research that would utilize several tests for assessing each of ten complex mental functions or “faculties”—memory, attention, imagination, comprehension, motor skill, and others. They

hoped thereby to reveal, in a brief period of time, the potential of such tests for the study of individual differences in specific intellectual capacities. The results of Stella E. Sharp's study (1899), some of the Binet-Henri tests, and his own subsequent researches convinced Binet that the study of specific faculties was a fundamental error—that what many of these tests tapped was a more general, unitary (albeit complex) intellectual process.

In their 1904 follow-up report on the results of their 1896 project, Binet and Henri reported that they too had failed to find any relatively economical or useful measure of individual differences. Binet had hoped that his ambitious program would reveal a test that, requiring an hour or two to administer, would reflect the individual differences in humans so obvious to the eye. After an eight-year search, as Charles SPEARMAN and a colleague wrote following Henri's presentation of this joint 1904 paper, Binet and Henri had failed to find a single test and could, for the present, only recommend the continued, long, systematic investigation of each individual being assessed (Wolf, 1969, p. 113). The experience that Binet gained from his controlled experimental laboratory studies would, however, be reflected in the 1905 Binet-Simon Scale.

THÉOPHILE SIMON AND ALFRED BINET

Almost one year to the day later, another chapter in the breakthrough and solution to the riddle came about, in part because of Binet's other interests and involvements. In 1899, a 26-year-old physician, Théophile Simon, came to the laboratory of the 42-year-old Binet, and asked to work with him (Wolf, 1961, p. 245). Simon was an intern of a Dr. Blin, at Perray-Vaucluse, an institution for mentally retarded and abnormal (then called “morally degenerate”) children and adults. After close scrutiny of the depth of Simon's motivation, Binet accepted him in 1899 and thereby began to forge another link in the chain of discovery. Simon was a devoted pupil and, later, a colleague; his vigor helped in the data collection from hundreds of subjects for their joint and more ambitious 1908 and 1911 revisions of the 1905 Binet-Simon Scale. Simon's, and later Binet's, work on the wards of Vaucluse pro-

vided Binet with access to the bona fide mental retardates he and Simon would use in 1905 as the lowest anchor points around which to compare the performance of normal and subnormal children in the Paris schools. This association with Simon probably also provided Binet with another point of contact with Dr. Blin—his student H. Damaye, who was studying the intellectual processes of mental retardates under Blin's direction at Perray-Vaucluse. As will be described shortly, Blin and Damaye, in 1902, would provide another link in the chain. Simon was also carrying out research under Blin's direction and, in 1900, Simon published a thesis on an anthropometric study of the head measurements of 223 retarded boys, for his M.D. degree (Wolf, 1969, p. 126). Simon's thesis impressed Blin so much that he accorded Simon a second year of internship to continue his studies of retardates; he also encouraged Simon to study with Binet, who had developed more exact methods for such measurements.

During the next few years, Binet and Simon published some dozen papers related to cephalometry, (head measurements), including norms for age levels and a search for differences between normals and retardates. Binet's association with Simon brought him into contact with the publications of Blin (1902) and the M.D. thesis of his student Damaye (1903); it reported on twenty test items that Damaye and Blin had developed as a crude test of *global* intelligence for more objectively differentiating the three clinically recognized forms of mental deficiency—idiot, imbecile, and moron. Until then, the diagnosis differentiating mental retardation from normal as well as the further classification of the three grades of mental retardation then in vogue were subjective; therefore, diagnoses differed among examiners or even by one examiner on repeated examinations of the same person. Binet and Simon, in the introduction to their 1905 scale, acknowledged that the Blin-Damaye test, a twenty-item oral questionnaire, was "superior to anything previously accomplished." The twenty test items (presented below) contained forerunners of the type of questions that would later appear in the Binet-Simon Scale. Binet's major criticisms were that the Blin-Damaye test (1) contained questions that were "superfluous," or could be answered merely by "yes" or "no," requiring little thought or judgment; (2) employed a system (0 to 5 points) for scoring each test reply that

was too subjective; (3) failed to provide the responses of normal children as anchor points for comparison; and (4) most important, failed to provide what Binet and Simon later called an *age scale* or other form of gradation of intelligence, which would allow the examiner to ascertain at once how much behind (or ahead) a child was in intellectual development. Blin and Damaye's test yielded a total score with a crude range for normal without taking into account the all-important effect of age on test scores.

FACTORS LEADING TO BINET'S BREAKTHROUGH

Binet had studied hypnosis as a global characteristic of behavior, had studied the development of his two young daughters, and then shifted to a study of separate faculties in his 1896 program of research with Henri. Through his own subsequent research and that of the American Titchenerian, Stella Sharp, he had realized his failings. He then found himself returning to the earlier notions of H. Taine—that intelligence was a global property of performance and could not be separated into specific faculties. Blin and Damaye, two practitioners involved in everyday clinical work with patients, had now helped reorient his thinking along practical lines. Binet's break with the type of mental testing done by Charles Spearman, Edward L. THORNDIKE, James McKeen CATTELL, and others was taking shape. By 1903, he conceived of intelligence as a global process that perceives external stimuli, then organizes, chooses, directs, and adapts such stimuli; that individuals differed widely in this capacity. He still had not devised a method of adequately assessing these differences. In 1903 his thoughts, although not free of testing the separate faculties of intelligence, began to move away from this approach.

THE ROLE OF LA SOCIÉTÉ IN THE BREAKTHROUGH

Another crucial link in Binet's discovery of the concept of mental age, with a single scale for its assessment, came from his concurrent activities in a study group known as La Société. As Wolf (1969, p. 132) discerned from a reading of the organization's early minutes, membership in this Society for the Psychological Study of Children was open to any and all in-

terested in studying normal children. It soon came to include teachers in the elementary, secondary, and early college grades; principals and other school administrators; lawyers; judges; medical doctors; psychiatrists; psychologists; sociologists; parents; and others. The society was organized in 1899 by Ferdinand Buisson, professor of education at the Sorbonne; from the very beginning it was dominated by Binet.

The role that La Société played in Binet's breakthrough developed because of the opportunities it provided. As a scientist, he tested his ideas on children involved in real-life situations, where decisions concerning their level of intellectual functioning would profoundly affect their subsequent educational experiences and thus their future lives. His association in La Société with influential leaders in education, law, politics, medicine, psychology, pedagogy, and other relevant professions provided him an opportunity to form study groups ("commissions") that worked on a variety of interrelated problems. Members of La Société created (1) a commission on graphology (handwriting; 1901), devoted to its general study but later including its potential to separate the retarded from the normal child (1903); (2) a commission for the study of the retarded (1904); and (3) a commission on memory (1904). There were others, including palmistry, by which he hoped to distinguish normal from subnormal.

At that time, intense social pressure existed in Paris for the separation of children who were fully educable from those educable with special help in the schools or those retarded to the point of being unable to benefit from public-school instruction. The sixteen members of the society's commission for the study of the retarded proposed at the February 1904 meeting that the society insist that (1) a medico-pedagogical examination be authorized by the school authorities before a child was denied public instruction due to mental retardation; (2) those children diagnosed educably retarded be educated in a special class or special establishment; and (3) as a demonstration project, a special class for the educably retarded be opened in one of the public schools near the Salpêtrière.

Wolf (1969, p. 210) reports that this resolution was adopted unanimously by La Société, and three members were then appointed to take it as a proposal from the society to the ministry of public instruction. Binet

asked the society's commission for the study of the retarded to draw up a vast plan of research to establish scientifically, as well as to measure objectively, the differences—mental and anthropometric—that separate the normal from the abnormal child.

CHRONOLOGICAL AGE NORMS ARE AN ESSENTIAL INGREDIENT

Under Binet's guidance, the commission on memory had been studying in a classroom of one of its teacher-members, M. Parison. They assessed the relationship of a pupil's score on several objective tests of memory to intelligence as subjectively rated by all the previous teachers of this same pupil. The results, published in the July 1904 *Bulletin*, revealed a positive relationship between children's memory ability and teachers' ratings of their intelligence. Binet commended the research and added "one could ignore the teachers' judgments . . . and compare the children of *the same ages* who are in *different grades*," an idea he acknowledged had been proposed earlier by M. C. Schuyten of Antwerp, Belgium. Binet had read about the results of this particular memory research and had already asked some teachers to carry out this same experiment in two school grades—fifth and seventh—but on 12-year-olds in those grades. The results revealed vast differences between those youngsters. The mean memory score for 12-year-olds in the seventh grade was twice that of 12-year-olds in the fifth grade. Binet realized that these results had potential for creating an objective index of individual intellectual differences.

In the November 1904 issue of the *Bulletin*, Binet announced the appointment, by the minister of public instruction, of a ministerial commission for the abnormal (i.e., a commission for the study of retarded children). The members included Binet and three other members of La Société. Binet's appointment to this commission—with its responsibility for evolving an objective method for the diagnosis of different types of retardation—provided the social "necessity" for which, in a short eight months (June 1905), the Binet-Simon Scale would become the "invention."

As late as January 1905, Binet had still not achieved the insight that led him and Simon to the first scale. V. Vanev was the principal of a school in which Binet

set up a pedagogical clinic in 1905. Under Binet's guidance, Vaney developed the world's first age-related achievement test for arithmetic; it could objectively grade academic retardation in terms of one, two, or three grades, relative to a child's age peers. For example, Vaney's research led him to report that at age 7, first grade, a child could read and write from dictation the numbers 1 to 20; also add and subtract among them orally. At age 8, second grade, a child could do the same up to 100; multiply any number from 1 to 10 by 2, 3, 4, and 5; and divide numbers from 1 to 20 by 2, 3, 4, and 5. Vaney proceeded through ages 9, 10, 11, 12, and 13, which was seventh grade, citing similar objective criteria. As early as 1845 in the United States, Horace Mann had introduced a standardized written test to replace the subject-by-subject oral examination of his day. In 1897, Rice followed this with a standardized written vocabulary test consisting of fifty items. E. L. Thorndike would soon introduce numerous other standardized written achievement tests. Yet Vaney's test is important because, by its age-related benchmarks, it provided a way to measure age-normed individual differences in French schools.

Binet was impressed, but he described Vaney's work only as a contribution to pedagogy that La Société had inspired. From Wolf's account of these several attempts at assessing intelligence, it is apparent that Vaney's work on age differences in performance helped Binet see a tie to the earlier crude Blin-Damaye global scale of twenty items. The latter included the following:

1. What is your name?
2. When were you born?
3. Are your parents living?
4. What do they do?
5. Put out your tongue.
6. Put your finger in your left eye.
7. Go to the wall and come back here.
8. Experiment with little dots.
9. Name objects shown—key, pen, pencil, etc.
10. What color is this pencil?
11. Are you less thirsty when it is hot than when it is cold?
12. What time is it?
13. Is a week longer than a month?

14. Where are you, here?
15. Is Brittany in France?
16. What do soldiers have on their heads?
- 17, 18, 19. Questions on reading, writing, spelling, and arithmetic.
20. What is the difference between the Catholic religion and the Protestant religion? (Varon, 1936, p. 43).

THE 1905 BINET-SIMON SCALE

Binet and Simon objected to this Blin-Damaye test; they found some questions to be superfluous, that a subjective method of scoring responses was employed, and that no normal children were used as anchor points—thus providing no opportunity to assess relative ability. Their objections are contained in their introduction to the 1905 Binet-Simon Scale. They also introduced their scale as superior to the Blin-Damaye; it contained the following thirty tests *arranged in ascending order of difficulty*:

1. Following a moving object with one eye.
2. Grasping a small object that is touched.
3. Grasping a small object that is seen.
4. Recognizing the difference between a square of chocolate and a square of wood.
5. Finding and eating a square of chocolate wrapped in paper.
6. Executing simple commands and imitating simple gestures.
7. Pointing to familiar named objects.
8. Pointing to objects represented in pictures.
9. Naming objects in pictures.
10. Comparing two lines of markedly unequal length.
11. Repeating three spoken digits.
12. Comparing two weights.
13. Showing a susceptibility to suggestion.
14. Defining common words by function.
15. Repeating a sentence of fifteen words.
16. Telling how two common objects are different.
17. Retaining a memory of a picture.
18. Drawing a design from memory.
19. Repeating a longer series of digits than the three in item 11.
20. Telling how two common objects are alike ("similarities").

21. Comparing two lines of unequal length.
22. Placing five (blocks) weights in order.
23. Designating which of the five weights has been removed.
24. Making rhymes (e.g., “What rhymes with ____?”).
25. Completing sentences.
26. Using three proffered nouns in one sentence.
27. Replying to twenty-five abstract (comprehension) questions (e.g., “When a person has offended you, and comes to offer apologies, what should you do?”).
28. Reversing the hands of a clock (in telling time).
29. Folding and cutting paper.
30. Defining abstract terms (e.g., What is the difference between esteem and friendship? . . . boredom and weariness? . . . etc.).

The ascending order of difficulty of these thirty tests in the 1905 scale was ascertained empirically by Binet and Simon, who based the order on results from their original standardization group of 100 (later reduced to 50) normal children, aged 3 to 12 years, and an unspecified number of mentally retarded children. The 1905 scale was developed to sample a wide range of functions, especially judgment, comprehension, and reasoning, which Binet felt constituted the essence of intelligence.

THE 1908 SCALE AND THE CONCEPT OF MENTAL AGE

The 1905 scale was published by Binet and Simon as a preliminary and tentative instrument for sampling intellectual behavior, not as a finished product. Their 1905 report contained no precise or objective method for arriving at either a score or an index. Rather, this first scale was meant to provide an approximation of the level of each child’s intellectual development. Binet and Simon continued their development of this crude “Measuring Scale of Intelligence,” as they called it, in their 1908 report on the further progress of their work. Not until the 1908 report did they formally introduce the concept of *mental age*, by specifically listing the three to eight items that could be passed by a majority of children at each age level from 3 through 12 years. Additionally, in this 1908 scale, the number of items had been increased from thirty to fifty-eight;

some of the earlier items had been discarded and others added; and, as just noted, the items were grouped into clusters for different age levels. With the introduction of age levels for different items, the revised 1908 Binet-Simon test permitted an examiner to judge, in units of one year, any given child’s mental age. Thus had the first objective, practical measure of intellectual functioning come into being.

MEASURED INTELLIGENCE AND ASSESSMENT

Binet and Simon were sensitive practitioners and they were theoretically oriented scientists. This is clear from the care with which they described the essential features of the psychological assessment of a child in 1905 and 1908. Their description of the “general conditions of the examination” was:

First the testing should take place in a quite isolated room. The examiner should be alone with the child . . . [who] should be kindly received; if he seems timid he should be reassured at once, not only by a kind tone but also by giving him first the tests which seem most like play, for example—giving change for 20 sous. Constantly encourage him during the tests in a gentle voice; one should show satisfaction with his answers whatever they may be. One should never criticize nor lose time by attempting to teach him the test; there is a time for everything. The child is here that his mental capacity may be judged, not that he may be instructed. Never help him by a supplementary explanation which may suggest the answer. Often one is tempted to do so, but it is wrong.

Do not become over anxious nor ask the child if he has understood, a useless scruple since the test is such that he ought to understand. Therefore one should adhere rigorously to the formulas of the experiment, without any addition or omission. Encouragement should be in the tone of voice or in meaningless words, which serve only to arouse him. “Come now! Very good! Hurry a little! Good! Very good! Perfect! Splendid! etc., etc.” If witnesses are inevitable, impose on them a rigorous silence.

Always begin with the tests that fit the child’s age. If one gives him too difficult work at first, he is discouraged. If, on the contrary, it is too easy it arouses his

contempt, and he asks himself if he is not being made fun of, and so makes no effort. We have seen examples of this misplaced self-esteem (Binet and Simon, 1908).

Binet and Simon insisted in the paragraph that followed those quoted above that the psychologist, while carrying out the standardized examination, should *not* be influenced by information about the child obtained from other sources; rather, that the psychologist should obtain a thorough knowledge of each child through testing. Binet and Simon saw a complete *assessment* as consisting of a tripartite approach to a child: (1) psychological; (2) pedagogical; and (3) medical.

In the general instructions for administering their scale, they insisted that the psychometric aspect of psychological assessment be objective—not influenced at the moment of its conduct by information from other sources. The practicing psychologist, physician, or educator-administrator could then make allowances when all the information was *integrated* to form a sound decision—a view recently rediscovered.

BINET'S DEATH, WORLD WAR I, AND RELIANCE ON IQ ALONE

Binet saw his test as a technique for *sampling* a child's current intellectual behavior and not as a rigid, fully developed, finalized test for all time of an individual's (innate) intelligence. He argued against finer gradations of mental age than a whole year (e.g., 6, not 6.5), resisting the finer gradations into tenths of a year that would come later. Although Binet died in 1911, before Wilhelm Stern (1912, translated by Guy M. Whipple, 1914) replaced the concept of mental age by intelligence quotient, it is fair to assume from Simon's 1959 oral report (Wolf, 1961, p. 245) that Binet might have vigorously objected to the later developments that fixed a single unchanging IQ on the millions of children subsequently examined by variants of his method. In their 1905 publication, Binet and Simon reported that they already were working on the age levels that they published in the 1908 revision. In 1905, they did not wish to calibrate their scale into an instrument that would do anything other than *classify* the three degrees of mental retardation—idiot, imbecile, and moron. In examining their fifty normal children as a comparison group, however, it became

apparent that their scale could classify children of normal and above normal mental functioning as well. This change in emphasis from a study of mental retardation for the Paris school system in 1905 to a study of general intellectual functioning among schoolchildren by 1908 is made clear in the titles of the two papers: "New Methods for the Diagnosis of the Intellectual Level of Subnormals" (1905) as opposed to "The Development of Intelligence in the Child" (1908). Had he lived, the probable direction Binet's work would have followed is clear from the title of his second and last revision: "New Investigation on the Measures of the Intellectual Level among Schoolchildren" (1911).

By 1911, Binet began to foresee numerous uses for his method in the study of child development, in education and medicine, and in longitudinal studies predicting differing occupational histories for children of differing intellectual potential. The 1911 revision was thorough. It involved equalizing the number of items at 5 for each age level—but extending the age levels upward, to include 15-year-olds, adding five tests for adults (ungraded), and relocating many of the test items or questions. By 1911, the two earlier scales had been translated into many languages. Despite a certain crudity, the scales were proving highly useful (valid) clinically in the hands of practitioners like H. H. Goddard in the United States (who unfortunately saw intelligence as a fixed, innate faculty) and among numerous European practitioners.

In reading Binet (Wolf, 1969, pp. 226–230), one concludes that he was interested in intellectual functioning and its potential for change as based on adaptive behavior. Binet would have resisted vigorously the heredity–environment controversy of the next research generation, considering it a pseudo-problem, born of an incomplete understanding of the nature of psychosocial assessment, on the one hand, and the crudity of his early test forms (tools), on the other. In 1911, Binet's premature death at age 54 deprived us of the full explication of his position. It became possible to develop objective empirical indices of "intelligence." With the 1916 revision developed by Stanford University's Lewis TERMAN, the Stanford-Binet, as the new test came to be known, would soon be translated into many languages and used worldwide by school and clinical psychologists before even more refined

testing methodologies and interpretive procedures were devised.

(See also: MENTAL AGE; STANFORD-BINET INTELLIGENCE SCALE.)

BIBLIOGRAPHY

- BINET, A. (1911). Nouvelles recherches sur la mesure du niveau intellectuel chez les enfants d'école. *L'Année Psychologique*, 17, 145–201.
- BINET, A., & SIMON, T. (1905). (1) Sur la nécessité d'établir un diagnostic scientifique des états inférieurs de l'intelligence. (2) Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. (3) Application des méthodes nouvelles au diagnostic du niveau intellectuel chez des enfants normaux et anormaux d'hospice et d'école primaire. *L'Année Psychologique*, 11, 162–336.
- BINET, A., & SIMON, T. (1908). Le développement de l'intelligence chez les enfants. *L'Année Psychologique*, 14, 1–94.
- BINET, A., & SIMON, T. (1916). *The development of intelligence in children. (The Binet-Simon Scale)*. Translated by E. S. Kite. Baltimore: Williams & Wilkins.
- MATARAZZO, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th and enl. ed.). New York: Oxford University Press.
- POLLACK, R. H. (1971). Binet on perceptual-cognitive development or Piaget-come-lately. *Journal of the History of the Behavioral Sciences*, 7, 370–374.
- POLLACK, R. H., & BRENNER, M. W. (Eds.). (1969). *The experimental psychology of Alfred Binet: Selected papers*. New York: Springer.
- STERN, W. L. (1912). Über die psychologischen Methoden der Intelligenzprüfung. *Ber. V. Kongress Exp. Psychol.* 16, 1–160. (English translation by G. M. Whipple [1914]. The psychological methods of testing intelligence. *Educational Psychology Monographs*, 13, Baltimore: Warwick & York.)
- TERMAN, L. M. (1916). *The measurement of intelligence*. Boston: Houghton Mifflin.
- VARON, E. J. (1935). The development of Alfred Binet's psychology. *Psychological Monographs*, 46 (3, Whole no. 207).
- VARON, E. J. (1936). Alfred Binet's concept of intelligence. *Psychological Review*, 43, 32–58.
- WOLF, T. H. (1961). An individual who made a difference. *American Psychologist*, 16, 245–248.
- WOLF, T. H. (1964). Alfred Binet: A time of crisis. *American Psychologist*, 19, 762–771.
- WOLF, T. H. (1966). Intuition and experiment: Alfred Binet's first efforts in child psychology. *Journal of the History of the Behavioral Sciences*, 2, 233–239.
- WOLF, T. H. (1969). The emergence of Binet's conceptions and measurement of intelligence: A case history of the creative process. Part II. *Journal of the History of the Behavioral Sciences*, 5, 207–237.

JOSEPH D. MATARAZZO

BIOECOLOGICAL THEORY OF INTELLECTUAL DEVELOPMENT

The bioecological theory of intellectual development proposes to account for the unevenness of intellectual performance in different contexts (for example, the rocket scientist who cannot grasp the gist of a literary passage; the street urchin who accurately makes change in a half dozen foreign currencies but cannot do a simple arithmetic problem; and the blurrier case of the student who scores 98 on the Miller Analogy Test but cannot negotiate a contract or assemble a VCR). It also proposes to avoid the limitations of reified unitary measures of intelligence, such as IQ, in predicting complex behavior in the everyday world (for instance, the gambler who exhibits a high degree of complexity but scores low on IQ). The three main concepts of the bioecological theory are these:

1. There are multiple innate abilities, as opposed to a single ability.
2. Innate abilities serve as a range of possibilities that develop, fail to develop, or develop, but later atrophy.
3. Motivation is important.

Like psychometric theories, the bioecological theory holds that intellectual growth is limited by innate capacity. But unlike psychometric theories, it holds that there is not just one capacity but many. So it is possible to be well endowed for some intellectual feats but not others. Second, in contrast to psychometric theories' view of innate intellectual capacity as pretty much fixed at birth and stable thereafter, the bioecol-

ogical theory holds that each innate ability is a *potential* ability, which will develop, fail to develop, or develop and sometimes atrophy as a result of its interaction with various aspects of the environment. Finally, the bioecological theory holds that motivation determines which innate potential abilities get developed into actual ones. So it is not enough to possess a potential innate ability or even possess it and also live in an environment that favors its development and expression; one must also be motivated to develop it.

Bioecological theorists refer to innate abilities (or, to be more precise, to innate potential abilities) as *biological resources* or *biological resource pools*; the total environment with which a person interacts is referred to as the *ecology*; and the set of specific aspects of the environment that makes possible maximum development of a specific ability is referred to as an *ecological niche*. As in other theories, the term *genotype* means the total genetic endowment of the individual, and the term *phenotype* means the physical expression of that portion of the genotype that gets actualized (not all of the genotype gets actualized).

Tenet #1: Multiple Abilities. The bioecological theory is, as its name implies, a blend of biological and ecological factors. The biological part of this theory holds that human intelligence is comprised of a system of biological resource pools, each responsible for a different aspect of information processing (e.g., visual rotation, contrast detection, angle displacement). These multiple biological sources of intellectual ability differ from those specified by traditional psychometric theories, because, unlike the latter, the bioecological theory denies that so-called general intelligence (or g) reflects a single biological resource that permeates all intellectual feats. This requires some explanation, given the venerable status that the concept of general intelligence (g) has occupied throughout this century.

The evidence for the existence of g comes from factor-analytic studies: simply put, whenever a battery of tasks are factor analyzed, a common ingredient can be shown to exist that all of the tasks seem to share, over and above their unique ingredients. So, for example, while a vocabulary test may reflect a set of specific skills (e.g., verbal fluency, skill at inferring meaning from context), it also is said to reflect the

operation of some common or general intellectual resource, known as g .

According to the bioecological theory, such a common factor can result from overlapping specific ingredients (that possess similar intercorrelations with each other) rather than a shared single ingredient that is common to all tasks. The factorist Godfrey THOMSON (1948) proposed the related view that multiple shared neural bonds rather than a single bond could underlie commonalities in a factor structure. For example, suppose that a battery of three tests (arithmetic, vocabulary, and spatial reasoning) is administered to children and that their scores are factor analyzed and g is extracted, indicating that performance on all of these tests is intercorrelated, with the child who is poor at one being generally poor at the others. Now suppose that performance on the arithmetic test depends on three cognitive ingredients: a (verbal encoding or the interpretation of words), b (a spatial mapping skill that is relevant for geometric problems), and c (a highly specific quantitative skill that is useful for a broad array of arithmetic problems, but is not useful for non-arithmetical problems). Also suppose that the two other tests (vocabulary and spatial reasoning) also sample some, but not all, of these same ingredients, in addition to some ingredients that are highly specific to each of them. For example, suppose that vocabulary requires a , verbal encoding, as well as d , the ability to "compare representations," and e , a highly specific vocabulary skill. Finally, suppose that spatial reasoning requires b , in order to engage in spatial mapping; d , in order to compare representations; and f , a highly specific spatial skill. Then vocabulary and arithmetic performances might be correlated because they share a verbal encoding ingredient (a), arithmetic and spatial reasoning might be correlated because they share a mapping ingredient (b), and vocabulary and spatial reasoning might be correlated because they share the ability to compare representations (d). Accordingly, g could end up being substantial in magnitude without actually representing a single ingredient that is common to all three tasks. In other words, performance on one test would be correlated with performance on the others not because all of these tests are saturated with some general intellectual resource but because they are multicomponent tests that involve some con-

stellation of partially overlapping ingredients. Children who are raised in an enriched ecology (environment) that fosters spatial skills (e.g., a home or school with plentiful supplies of jigsaw puzzles), tend also to have enriched environments that foster verbal and quantitative skills. Thus, g is seen as the result of correlated environments, not the result of a single biological resource that imbues every type of intellectual test.

Tenet #2: Interaction and Synergy. The bioecological theory is inherently developmental and interactionist, holding that biological potentials are developed and shaped by the environment, while at the same time they help to shape the environment. Like all interactionist perspectives, the bioecological theory asserts that from the very beginning of life there is an interplay between biological potentials and environmental forces. In order to understand how individuals can begin life possessing comparable intellectual potentials but differ in the level of intelligence they subsequently manifest, the bioecological theory posits an interaction between various biologically influenced resources (innate potential abilities), such as the capacity to store, rotate, and scan information, and the ecological contexts (particular sets of environmental features) that are relevant for the unfolding of each potential ability. At each point in development, the interplay between biology and ecology results in changes that may themselves produce other changes until a full cascading of effects is set in motion.

Although biology and ecology are interwoven in an indivisible whole, their relationship is continually changing, and with new change a new set of possibilities is set in motion until soon even small changes produce cascading effects. Hence intellectual change is not always or even usually linear, but rather is synergistic and nonadditive. A small environmental influence on a protein-fixing gene may initially result in only tiny intellectual changes, but over time the chain of events may produce a magnification of effects on other processes. In addition, certain periods in development can be thought of as sensitive periods during which a unique disposition exists for a specific intellectual "muscle" to develop in response to its interaction with the environment. During such periods, neurons within specific compartments rapidly "overarborize" (spreading their tentacle-like synaptic

connections to other neurons in response to environmental stimulation). Although some of the arboreal connections laid down during these periods of brain spurts will not be used at the time, they can be recruited to enable future behaviors to occur, provided they are not "pruned" because of atrophy or disuse. It appears that while some neural processes are more fully under maturational control, others are much more responsive to the environment, and synapses are formed in response to environmental stimulation (learning) that may vary widely among humans.

Tenet #3: Motivation. The bioecological theory incorporates *motivation* as a key ingredient in its explanation of empirical findings. Briefly, it is not sufficient for an individual to be endowed with some biological potential for a given intellectual ability, or merely to be exposed to an environment that is important for the expression of this ability; the individual must also be motivated to benefit from exposure to such an environment. The gamblers in S. J. Ceci and J. Liker's study (1986) who demonstrated highly complex forms of reasoning at the racetrack did not exhibit the same degree of complex reasoning in other domains. Had they been exposed to environments that were conducive to, say, learning science or philosophy, and motivated to take advantage of such environments, they undoubtedly would have acquired the ability to think as complexly in those domains as they did at the racetrack, given the similarity between the type of reasoning needed to handicap a race and that needed to reason scientifically.

The Mismatch Between g and Real-World Intelligence. Since the three tenets of the bioecological theory of intellectual development have been described, it now can be contrasted with other views of intelligence on five grounds.

First, the bioecological view, like other theories that posit the existence of multiple intelligences (e.g., Gardner, 1983; Sternberg, 1985), proposes that these multiple cognitive abilities are, at best, only imperfectly measured by tests of so-called general intelligence (IQ). For example, Ceci (1990) provides numerous examples of IQ failing to predict cognitive complexity in the workplace. According to the bioecological theory, IQ is seen as a marker for a specific set of verbal/academic skills that are acquired in school

and in school-related activities (e.g., plays, class trips). In this sense, IQ is an achievement test much like those that routinely are administered to students in school. It is susceptible to the same influences as all achievement tests, namely, the time spent studying the material, the motivation to benefit from instruction, and the biological capacities that are relevant for the type of learning involved. To see this clearly, one needs only look at the relationship between the amount of time one spends in school and one's IQ (as well as math, reading, and science achievement scores). Basically, the evidence is quite strong that staying in school props up one's IQ and achievement scores. Once schooling is terminated, children's IQs begin a steady decline. There is nothing mysterious about this if one views IQ as a form of achievement that is learned in school. This does not deny that biology is involved, too, because biology is involved in virtually everything that is cognitive. But it does depart from the simplistic notion that IQ scores are direct reflections of the biological capacity to learn. Instead, it suggests that IQ is itself influenced by learning.

Second, the bioecological theory differs from traditional nature–nurture interactionist views in the conception of the interaction between biological and environmental factors. The traditional interactionist view gives primacy to the importance of biological factors (“genes encapsulate phenotypes,” meaning that the physical expressions of our genetic endowment are fully under biological control, as opposed to environmental control). The bioecological theory departs from this deterministic view of the genotype and gives the environment a more significant role. Genes do not encapsulate phenotypes; rather, they manufacture proteins and enzymes that influence the expression of neighboring genes as well as interact among themselves. This implies that such interactions are governed by physical and chemical laws independent of the strand of DNA from which they originated (Subtelny & Green, 1982). A model of the translation of genotypes into phenotypes requires that we consider not just the proteins that genes manufacture but also the developmental role such proteins play, since most of the hormones, inducers, and inhibitors are connected in complex ways with the activity of multiple gene systems. Thus the resultant brain structures are only indirectly related to genes, making it impossible to ex-

plain their development exclusively or even primarily in terms of genes.

Third, like other interactionist views of development, the bioecological theory argues that the efficiency of cognitive processes depends on aspects of the context. However, according to Ceci (1990), context is not an adjunct to cognition, but a constituent of it. That is, unlike traditional theories of intelligence, which assume that context is merely a background for cognition, the bioecological theory regards context as an inextricable aspect of intellectual efficiency. Here context is defined broadly to include not only external features of the near and far environments and their motivational properties, but internal features of the organism's mental representation, such as the manner in which a stimulus or problem is represented in memory. Thus, speed in recognizing letters and numerals depends on how those stimuli are represented in memory, with more elaborate representations leading to faster recognition rates. This explains why the same cognitive ability, no matter how basic, operates inconsistently across diverse contexts. The same individuals who are slow at recognizing a word in one domain (e.g., sports), may recognize it in another domain (e.g., culinary terms) more quickly if its representation in the latter domain is more elaborate. In short, intelligence-in-context research has shown that context, including the mental representation or mental context of a task, determines the efficiency of intelligence.

Fourth, the bioecological theory assumes that there exists many nonintellectual abilities that are highly important for subsequent intellectual development but are nevertheless inherited. For example, a child may inherit various types of temperament (e.g., restlessness, impulsivity), physical traits (e.g., skin color, facial shape), and “instigative characteristics” (e.g., “reward-seeking”) that may influence later learning and development. While these traits are themselves influenced by gene systems, and can be shown to exert direct as well as indirect effects on subsequent IQ performance and school success, they are not intellectual in nature. So, these nonintellectual characteristics and abilities can account for part of the heritability pattern (e.g., IQs that run along family lines) without claiming that this is a consequence of the inheritance of a central nervous system with, say, a set signal-to-noise ratio that limits processing capacity. Family members who

share such characteristics (e.g., impulsivity) may perform similarly as a result of these nonintellectual dispositions, rather than because they share the same rate-limiting nervous systems. Significantly, accounts of rate-limiting intellectual functioning that are based on EEG-power spectral-density measures, blood glucose levels in the brain, central nerve-conductance velocity and oscillation, and heritability analyses cannot distinguish between intellectual (i.e., inherited limitations on one's central nervous system functioning) and nonintellectual bases of performance that are biologically based (e.g., impulsivity). Again, this is not to deny the importance of genes in intellectual performance, but merely to caution that just as it is the case that not all that is intellectual is genetic in origin, not all that is genetic is intellectual in nature.

Finally, the bioecological theory departs from traditional behavior genetic models regarding the nature and meaning of heritability, or h^2 . The latter hold that intellectual ability is transmitted according to the degree of consanguinity, with children sharing approximately 50 percent of their parents' genes, and therefore resembling their parents' IQs by a similar magnitude. According to the bioecological theory, h^2 reflects the proportion of "actualized" genetic potential, leaving unknown and unknowable the amount of unactualized genetic potential. Thus nothing can be said about an individual's biological potential for success or failure without some information about the level of environmental resources and motivation that exist in the child's environment to translate various genetic potential abilities into various phenotypic or expressed abilities. If there are insufficient resources and motivators in one's life, then h^2 will reflect only that portion of one's potential that can be brought to fruition by the limited resources available.

In sum, the bioecological theory is an attempt to craft an account of intellectual functioning that is mindful of the developmental, cultural, and psychometric findings. To encompass all of these factors, the theory has abandoned reliance on notions of general intelligence that are the result of a single biological resource (e.g., signal-to-noise ratio in the transmission of information in the nervous system). It has elevated the role of context from one of background noise to one of centrality, and it has incorporated motivation into its tenets. To undertake some intellectual feat at

a high level, it is not sufficient to be endowed with a specific biological resource; one must also be exposed to the relevant educational experiences, and be motivated to benefit from them.

(See also: INTERACTIONIST VIEWS ON INTELLIGENCE; NATURE, NURTURE, AND DEVELOPMENT; THOMSON'S RANDOM OVERLAP THEORY.)

BIBLIOGRAPHY

- ALBERCH, P. (1983). Mapping genes to phenotypes, or the rules that generate form. *Evolution*, 37, 861–863.
- BRONFENBRENNER, U., & CECI, S. J. (1993). Heredity, environment, and the question "how"? A first approximation. In R. Plomin & G. McClearn (Eds.), *Nature, nurture, and psychology* (pp. 313–325). Washington, DC: American Psychological Association.
- CECI, S. J. (1990). *On intelligence . . . more or less: A bio-ecological treatise on intellectual development*. Englewood Cliffs, NJ: Prentice-Hall (Century Psychology Series).
- CECI, S. J., & LIKER, J. (1986). A day at the races: A study of IQ, expertise, and cognitive complexity. *Journal of Experimental Psychology: General*, 115, 255–266.
- GARDNER, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- STERNBERG, R. J. (1985). *Beyond IQ: A triarchic framework for intelligence*. New York: Cambridge University Press.
- SUBTELNY, S., & GREEN, P. B. (EDS.). (1982). *Developmental order: Its origin and regulation*. New York: Alan R. Liss, Inc.
- THOMSON, G. H. (1948). *The factorial analysis of human ability* (3rd ed.). Boston: Houghton-Mifflin.

STEPHEN J. CECI

BIOLOGICAL MEASURES OF INTELLIGENCE

In the late nineteenth century, Francis GALTON was engaged in state-of-the-art but premature efforts to discover sensory measures of intelligence. In 1905, Alfred BINET and Théophile Simon developed the first successful test for the measurement of intelligence, but none of its items measured sensory or biological functions. During the subsequent hundred years, others in the field attempted to add biological and other classes of items radically different from the primary paper-and-pencil items included in the Binet-

Simon Scale. Most such attempts were unsuccessful. Recent attempts utilizing very different approaches (i.e., sensory and other physiological indices) than those used by Binet and Simon, as well as their successors, however, have borne more fruit (reviewed in Eysenck, 1986, 1987, 1988; Kranzler & Jensen, 1989; Jensen, 1987, 1991; Vernon 1990; Vernon & Mori, 1990). Specifically, there is accumulating evidence from a number of laboratories that the scores of individuals who differ on traditional measures of Charles SPEARMAN's *g* (the presumed general factor underlying differences in intelligence scores), such as scores on the RAVEN PROGRESSIVE MATRICES, the Binet-Simon Scale, and the WECHSLER SCALES OF INTELLIGENCE, correlate statistically significantly with their own equally reliable "scores" on measures that also reflect sizable individual differences in a variety of biological indices. These biological indices include (1) individual differences in the average evoked potential (AEP) index obtained from electroencephalogram (EEG) recordings from the surface of the brain; (2) individual differences in the durations of reaction times (RTs) in responding to stimuli presented to the eye or ear, including both

single and more complex RTs (e.g., discrimination and choice RTs); (3) inspection times; and (4) the trial-to-trial, intraindividual oscillations and variabilities (i.e., standard deviations) in such person-specific physiological measures and reaction time measures. In addition, the results of early studies suggest that measures of both nerve conduction velocity and the rate at which glucose is metabolized in the brain also correlate with scores on these traditional measures of IQ.

Two examples of these biological correlates of the Wechsler Full Scale IQ are illustrative. The first is a study reported by D. E. Hendrickson (1982). The biological measure used by Hendrickson was the AEP obtained from each individual's own EEG recording. The EEG is a measure of the electrical activity of the brain that is recorded from electrodes attached to the skull. Figure 1 is a diagrammatic presentation of the resting EEG (to the left of the arrow), followed at point A in this case by an auditory (sensory) stimulus introduced by the investigator and reacted to cortically by the individual being tested as he or she processes the auditory stimulus just presented. Averaging the series of EEG waves following the stimulus, negative and then

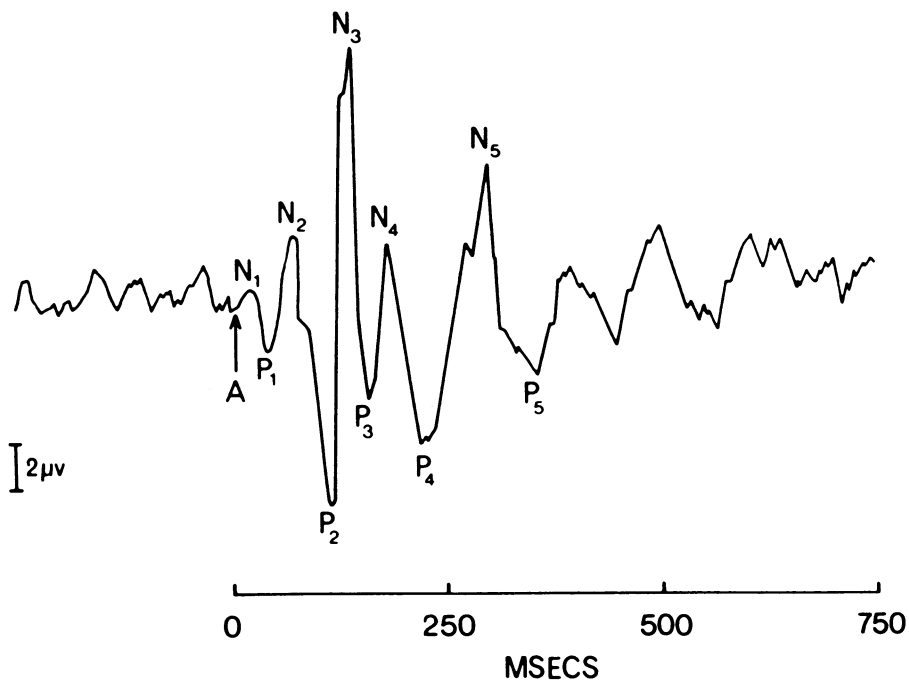


Figure 1

Following an auditory sensory signal. The stimulus was administered at point A.

SOURCE: Hendrickson, 1982.

positive in each case, produces the AEP for this individual, a wave process that gradually dies out after something like 750 to 1000 milliseconds. In Figure 1, successive negative and positive EEG waves are labeled N_1 , P_1 , N_2 , P_2 , etc., and are averaged to produce a biological index (AEP), which differs from person to person but which, although it is variable from trial to trial, nevertheless yields a reliable (stable), idiosyncratic index for each individual when averaged across successive presentation trials.

In his investigation, Hendrickson studied a sample of 219 older adolescents (121 boys and 98 girls), to each of whom he administered both a Wechsler Adult Intelligence Scale (WAIS) and an EEG, which produced the basis for the second measure (AEP). From the AEP, Hendrickson derived two additional scores: (1) the trial-to-trial variability (technically referred to as the standard deviation, or variance) of the AEP as each individual focused attention and processed the auditory information before executing a choice response; and (2) the complexity of the height, breadth,

number, and other physical dimensions of the AEP waves that the individual's brain produced while he or she was executing the choice-response cognitive task.

Hendrickson's remarkable findings are presented in Table 1. As is shown in the bottom row, Full Scale IQ was highly correlated with both of the two biological (EEG) measures recorded from the brain. The independently measured Full Scale IQ correlated $-.72$ with the EEG trial-to-trial variability (variance) measure and also correlated $.72$ with the complexity of the EEG wave measure derived from the AEP of the same individual. (What this means is that individuals who obtained higher IQ scores were also individuals whose EEG-derived brain waves [1] were the most complex and [2] varied in shape the least across successive presentations of the auditory choice-response task.) Equally striking are the correlations shown in the body of the table between these two biological measures and each of the eleven subtests of the WAIS measure of intelligence. In a study explicitly designed as an attempt independently to confirm Hendrickson's find-

TABLE 1
Correlations between WAIS subtest scores and AEP variance, complexity, and combined score

WAIS Subtest	Variance	Complexity	Complexity Minus Variance
Information	-.64	.55	.68
Comprehension	-.50	.53	.59
Arithmetic	-.57	.56	.65
Similarities	-.69	.54	.71
Digit Span	-.54	.49	.59
Vocabulary	-.57	.62	.68
Verbal (IQ)	-.69	.68	.78
Digit Symbol	-.28	.32	.35
Picture Completion	-.47	.52	.57
Block Design	-.50	.45	.54
Picture Arrangement	-.36	.45	.46
Object Assembly	-.32	.45	.44
Performance (IQ)	-.53	.53	.60
WAIS (Full Scale IQ)	-.72	.72	.83

SOURCE: Hendrickson, 1982.

ings, Stough, Nettelbeck, and Cooper (1990) also found high correlations between their AEP measures and the same individual's score on Verbal IQ (VIQ), Performance IQ (PIQ), and Full Scale IQ (FSIQ) as measured with the Revised WAIS (WAIS-R).

These latter findings, plus those of Hendrickson, appear to suggest that, as a measure of intelligence, each of these EEG measures is almost as robust a measure of the Spearman g loadings (i.e., a measure derived from a factor analysis of the WAIS subtests) as are the IQ indices derived from the Binet, Wechsler, and many group-administered paper-and-pencil tests of intelligence currently in use. Furthermore, in his review of this literature, Jensen (1987, pp. 105–106) underscores the results of a study by Eysenck and Barrett (1984), independently confirmed in a study by Schafer (1985), in which Eysenck and Barrett reported that the Spearman g loadings showed a rank-order correlation of .95 with the correlations of each of the twelve WAIS subtests with the AEP. In a study that independently repeated the 1982 Hendrickson study, however, Barrett and Eysenck (1992) were able to duplicate only the main thrust (but not all) of Hendrickson's findings—an indication that research on EEG and other biological correlates of IQ is still in its relatively early stages.

The second example is from a recent completed study by Reed and Jensen (1992), which involved an investigation of the relationship between an individual's IQ as measured by the Raven Advanced Progressive Matrices test and the same individual's independently determined average cerebral evoked potentials following the presentation of a visual stimulus. The investigators used 147 male undergraduates, from each of whom (in response to a visual pattern-reversal stimulation task) they obtained short-latency, visually evoked potentials (VEPs), comparable to those shown in Figure 1, which were recorded over the primary visual cortex at the back of the brain. Dividing each subject's head length by the average latency of his VEP provided a measure of each person's visual neural conduction velocity (NCV), as nerve impulses (of 70 to 100 msec) were transmitted from the retina through the brain's visual tract to the visual cortex. This nerve conduction velocity measure (V:P/100) for a group of VEPs averaging about 100 milliseconds showed a statistically significant correlation (.37) with each person's

Raven Progressive Matrices IQ score. To present the same findings diagrammatically, the distribution of the NCV measure from the 147 undergraduates, ranging from the lowest (1.75 m per sec), or least efficient, to the highest (2.22 m per sec), or most efficient, was divided into quintiles (i.e., each quintile contained the scores from 20 percent of the students). Averaging the Raven IQ scores for students in each nerve conduction velocity quintile, Reed and Jensen obtained the results shown here in Figure 2. Specifically, for the five groups of subjects with progressively and increasingly more efficient NCVs during the processing of the visually presented pattern-reversal information tasks, the average IQ scores were 114, 116, 117, 121, and 122, respectively. Whether these results are represented diagrammatically or more precisely by the reported Pearson product moment correlation of .37, they appear as remarkable as those of Hendrickson.

To add credence to the robustness of these Reed and Jensen findings, Vernon and Mori (1992), in a study of conduction velocity in the median nerve of the arm, found a correlation of .42 (and .48 in a confirmation study reported in the same paper) between Full Scale IQ and NCV. Although the correlations were lower in magnitude than those reported by Hendrickson, Vernon and Mori also found significant correlations between their arm NCV measures (as well as their reaction time measures) and scores of the individual subtests of their multiple-choice variant of the WAIS measure of intelligence. It is necessary to emphasize that research in this area is still in its beginning stages: Other investigators have not confirmed these Vernon and Mori findings. Thus, although Barrett, Daum, and Eysenck, (1990) found a statistically significant correlation of $-.44$ between the variability of the averaged sensory nerve action potentials in the hand and the Raven IQ score in a study of forty-four individuals, a study by Reed and Jensen (1991) failed to find a correlation between NCV in the arm and IQ score. The thrust of most of the findings is clear, however.

Some two decades ago, when correlations between IQ scores and biological measures were first being reported, most investigators in the field of intelligence were skeptical that the century-old search for biological measures of intelligence would bear fruit, and they remained so until recently. The increasing numbers of

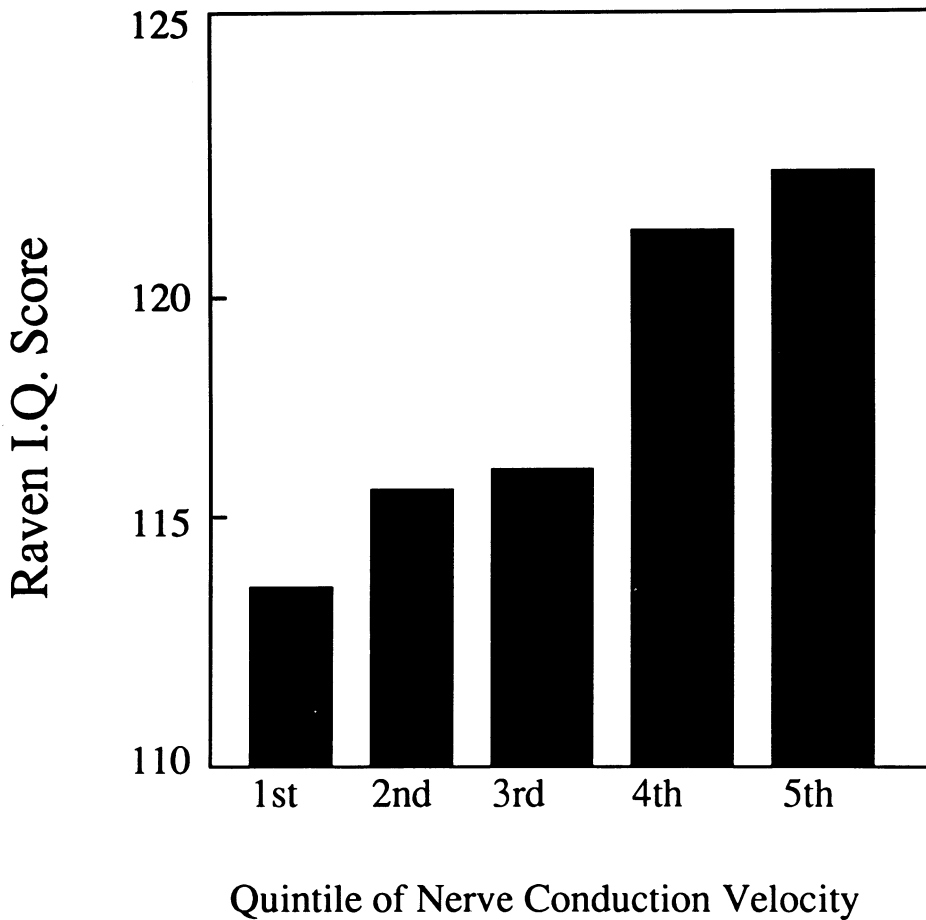


Figure 2

Mean Raven IQ score by quintile of nerve conduction velocity. The first quintile represents the slowest (and thus least efficient) velocities.

SOURCE: Reed and Jensen, 1992.

such studies, however, as well as the duplication of the positive findings in many—but, it is important to note, not all—laboratories in different parts of the world (see Kranzler & Jensen, 1989), has led a number of scientists to believe that in the foreseeable future some biological correlates will be shown to be valid measures of what up to now has been measured by traditional paper-and-pencil intelligence tests.

We should expect additional technological advances in both the biological measures that are recordable as well as the way they are recorded. It is reasonable to suppose that the technology for measuring these neurophysiological parameters of simple and more complex information-processing activities, as well as their

executed end-stage target behaviors, also will be improved considerably in the not-too-distant future. For example, one may anticipate that such improvements will come from the development of new generations of neurophysiological, neurochemical, and neuromolecular measures of information processing and related aspects of cortical functioning, as well as from insights yielded by new generations of advanced systems of brain-imaging techniques, such as successors to Positron Emission Tomography (PET). As an example, the types of technological measuring advances one may anticipate are discernible from a pilot study of eight volunteers that utilized PET by Haier et al. (1988). The study found significant negative correlations between

glucose metabolic rate (the rate at which sugar is used up) in the brain and a measure of abstract intelligence from the Raven Advanced Progressive Matrices test. This finding suggested that the individuals who obtained the best scores on the Raven intelligence test actually expended the least brain energy while responding to the items that comprise the test.

To test their hypothesis that people who obtain higher scores on IQ tests expend less brain energy, Haier et al. (1992) next measured the rate of glucose metabolism in the brain on two occasions. Glucose measurements were made during the initial trial, when their eight male subjects first began to learn (using the computer game Tetris) a complex video learning task, and again during the very last trial after four to eight weeks of daily practice of the complex video game skill. The results for the eight subjects indicated that over the four to eight weeks of practice, metabolic rate measured in surface regions of the brain decreased despite a more than sevenfold average improvement in game performance skill. Furthermore, the subjects who improved their performance the most after practice showed the largest glucose metabolic decreases in several areas of the brain. Haier and his colleagues interpreted this finding as suggesting that the learning of a complex motor skill response is associated with a decreased use of extraneous or inefficient brain areas. In a further test of their hypothesis, Haier, Siegel, Tang et al. (1992), correlated the scores earned on both the Wechsler Adult Intelligence Scale-Revised (WAIS-R) and on the Raven Advanced Progressive Matrices Intelligence test by each of their eight subjects with measures of rate of glucose metabolism in the brain of the same individual. That study yielded statistically significant findings to support their brain efficiency hypothesis. They found that individuals who score highest on measures of intellectual ability also expend lesser amounts of brain energy than do lower-scoring individuals. These IQ findings also supported their evolving belief that learning is correlated with the rate of glucose metabolism in some brain regions and not in others.

These findings by Haier and his colleagues are consistent with those reported from other laboratories. Specifically, a study by Parks et al. (1988), utilizing sixteen normal volunteers, and correlating a PET measure of glucose metabolic rate in the brain with the

individual's score on a neuropsychological test of verbal fluency, also yielded a negative correlation between glucose metabolic rate in the brain and verbal fluency as measured by the number of words beginning with different letters produced by the individual during each sixty-minute trial. Other studies that have reported significant correlations between rate of metabolism and measures of IQ are summarized by H. J. EYSENCK (1987, pp. 53-55), and many other studies reporting statistically significant correlations between the PET measure of metabolic rate and various types of intelligence measures are included in a literature review of this small body of studies by Haier (1992).

All the biological indices of intelligence discussed above are indices of brain function. It appears likely, however, that these indices will be complemented by measures of brain structure derived with future generations of modern imaging technologies. One example of such research was conducted by Willerman et al. (1991). After controlling for body size, these authors reported correlations between brain size and WAIS-R IQ of .65 in men ($N = 20$) .35 in women ($N = 20$), and .51 for both sexes combined. A further statistical correction for the sample of forty college students involved indicated a correlation for both sexes combined of about .35 in a more representative sample. A follow-up analysis of the same forty subjects by these authors (Willerman et al., 1992) suggested that in men a relatively larger left hemisphere predicted better WAIS-R subtest verbal than nonverbal ability, whereas in women a larger left hemisphere predicted relatively better nonverbal than verbal ability. Obviously, the small numbers of individuals used in this study requires the independent confirmation of these results by other investigators. Nevertheless, as increasingly refined neuroimaging technologies become available, other structural features of the brain in studies utilizing larger numbers of subjects and additional biological measures undoubtedly will become targets for assessment and may be found to correlate significantly with aspects of intellectual (if not also personality) functions.

At this point, a note of caution is necessary. The studies discussed above all reported positive results, and taken in toto, they suggest that valid biological measures of intelligence have been identified. Other experts in this field, however, believe that these studies

are too few in number and substance for anyone to draw the far-reaching conclusion that biological indices of intelligence have been identified. Psychologists and other cognitive scientists (although not all of them) have regarded these studies as questionable for three reasons:

1. The sizes of the samples studied are too small;
2. Some of the statistically significant correlations may be spurious and artifactual (i.e., the result of chance) due to the very large numbers of correlations computed in those studies rather than mirroring true biology-intelligence relationships; and
3. Most important, some investigators (who have not published their findings) who repeated one or another of the studies were unable to duplicate their own or others' positive findings.

Nevertheless, the number of positive findings in this area reported to date is of sufficient size that one should expect a larger number of studies of these types to be undertaken in the near future.

In summary, research on the biology of intelligence is, at most, only in the beginning stages. Given the potential of such research, however, it is possible to anticipate further development and use in practice of the measures described above and other biological indices of brain function and structure in a test (or a test battery) for the measurement of individual differences in intelligence and related aspects of mental ability—the first clear break from test items and tests in the Binet tradition in a century. New tests (or batteries of new tests) comprising only biological measures of cortical functioning may be developed that will predict, as well as do today's tests, success in school, as well as occupational attainment and other aspects of everyday living. Unless technological advances in miniaturization, as well as other inventions and developments relating to the practicability of administration (including social acceptability) also occur, however, it is unlikely that such biological tests will replace the group-administered tests, such as the Scholastic Aptitude Test used in schools and the comparable group tests used in industry. Nevertheless, if research in this area proves to be reliable and valid, the use of such new biological measures of ability by a school psychologist or a clinical neuropsychologist providing diagnostic

and rehabilitative help in the individual case (e.g., to a brain-injured adult or to a dyslexic or other learning-disabled child) is much more likely than is their larger-scale use with normal groups of children or adults.

(See also: BRAIN; EEG EVOKED POTENTIALS; MEASUREMENT AND PREDICTION OF INTELLIGENCE.)

BIBLIOGRAPHY

- BARRETT, P. T., DAUM, I., & EYSENCK, H. J. (1990). Sensory nerve conduction and intelligence: A methodological study. *Journal of Psychophysiology*, *4*, 1–13.
- BARRETT, P. T., & EYSENCK, H. J. (1992). Brain evoked potentials and intelligence: The Hendrickson Paradigm. *Intelligence*, *16*, 361–381.
- EYSENCK, H. J. (Ed.). (1982). *A model for intelligence*. New York: Springer Verlag.
- EYSENCK, H. J. (1986). The theory of intelligence and the psychophysiology of cognition. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 1–34). Hillsdale, NJ: Erlbaum.
- EYSENCK, H. J. (1987). Speed of information processing reaction time, and the theory of intelligence. In P. A. Vernon (Ed.), *Speed of information processing and intelligence* (pp. 21–67). Norwood, NJ: Ablex.
- EYSENCK, H. J. (1988). The concept of “intelligence”: Useful or useless? *Intelligence*, *12*, 1–16.
- EYSENCK, H. J. & BARRETT, P. (1984). Psychophysiology and the measurement of intelligence. In C. R. Reynolds, & V. Wilson (Eds.), *Methodological and statistical advances in the study of individual differences* (pp. 1–49). New York: Plenum.
- HAIER, R. J. (1992). Cerebral glucose metabolism and intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence*. Norwood, NJ: Ablex.
- HAIER, R. J., SIEGEL, B. V., MACLACHLAN, A., SODERLING, E., LOTTENBERG, S., & BUCHSBAUM, M. S. (1992). Regional glucose metabolic changes after learning a complex visuospatial-motor task: A Positron Emission Tomographic study. *Brain Research*, *570*, 134–143.
- HAIER, R. J., SIEGEL, B. V., JR., NUECHTERLEIN, K. H., HAZLET, E., WU, J. C., PAK, J., BROWNING, H. L., & BUCHSBAUM, M. S. (1988). Cortical glucose metabolic rate correlates of abstract reasoning and attention studied with Positron Emission Tomography. *Intelligence*, *12*, 199–217.

- HAIER, R. J., SIEGEL, B. V., TANG, C., ABEL, L., & BUCHSBAUM, M. S. (1992). Intelligence and changes in regional cerebral glucose metabolic rate following learning. *Intelligence, 16*, 415–426.
- HENDRICKSON, D. E. (1982). The biological basis of intelligence. In H. J. Eysenck (Ed.), *A model for intelligence* (pp. 151–228). New York: Springer Verlag.
- JENSEN, A. R. (1987). The *g* beyond factor analysis. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing* (pp. 87–142). Hillsdale, NJ: Erlbaum.
- JENSEN, A. R. (1991). General mental ability: From psychometrics to biology. *Diagnostique, 16*, 134–144.
- KRANZLER, J. H., & JENSEN, A. R. (1989). Inspection time and intelligence: A meta-analysis. *Intelligence, 13*, 329–347.
- PARKS, R. W., LOEWENSTEIN, D. A., DODRILL, K. L., BARKER, W. W., YOSHI, F., CHANG, J. Y., EMRAN, A., APICELLA, A., SHERAMATA, W. A., & DUARA, R. (1988). Cerebral metabolic effects of a verbal fluency test: A PET scan study. *Journal of Clinical and Experimental Neuropsychology, 10*, 565–575.
- REED, T. E., & JENSEN, A. R. (1991). Arm nerve conduction velocity (NCV), brain NCV, reaction time, and intelligence. *Intelligence, 15*, 33–47.
- REED, T. E., & JENSEN, A. R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence, 16*, 259–272.
- SCHAFFER, W. E. P. (1985). Neural adaptability: A biological determinant of *g* factor intelligence. *Behavioral and brain sciences, 8*, 240–241.
- STOUGH, C. K. K., NETTELBECK, T., & COOPER, C. J. (1990). Evoked brain potentials, string length and intelligence. *Personality and Individual Differences, 11*, 401–406.
- VERNON, P. A. (Ed.). (1987). *Speed of information processing and intelligence*. Norwood, NJ: Ablex.
- VERNON, P. A. (1990). The use of biological measures to estimate behavioral intelligence. *Educational Psychologist, 25*, 293–304.
- VERNON, P. A., & MORI, M. (1990). Physiological approaches to the assessment of intelligence. In C. R. Reynolds & R. W. Kamphaus (Eds.), *Handbook of psychological and educational assessment of children: Intelligence and achievement* (pp. 389–402). New York: Guilford Press.
- VERNON, P. A., & MORI, M. (1992). Intelligence, reaction times, and peripheral nerve conduction velocity. *Intelligence, 16*, 273–288.
- WILLERMAN, L., SCHULTZ, R., RUTLEDGE, J. N., & BIGLER, E. D. (1991). In vivo brain size and intelligence. *Intelligence, 15*, 223–228.
- WILLERMAN, L., SCHULTZ, R., RUTLEDGE, J. N., & BIGLER, E. D. (1992). Hemisphere size asymmetry predicts relative verbal and nonverbal intelligence differently in the sexes: An MRI study of structure-function relations. *Intelligence, 16*, 315–328.

JOSEPH D. MATARAZZO

BIRTH DEFECTS Teratology, which is a scientific subspecialty of biology, combines two words from the Greek: *terato* (monster) and *ology* (study). It is a word that has been retained, although infants born with abnormalities are not “monsters” in either colloquial or current medical usage. Teratology is the branch of embryology, pathology, and neuropsychology that examines physical malformations, abnormal development, and behavioral changes in offspring that occur as a result of an insult to the fetus during pregnancy. The term *fetus* will be used here to include the entire gestational period, although the term is technically appropriate only for the ninth week of pregnancy through delivery.

EXTENT OF THE PROBLEM

Some 250,000 U.S. infants, or 7 percent, are born annually with mental or physical defects. Of these, 20 percent, or 50,000, have genetic defects, such as Tay-Sachs disease (characterized by mental retardation and blindness) or cystic fibrosis (characterized by severe respiratory problems). Another 10 percent, or 25,000, are born with chromosomal abnormalities, such as DOWN SYNDROME, whose effects include altered body and facial features and mild to extreme mental retardation. Some inherited factors are sex linked. For example, Turner’s syndrome affects only females; these children have an XO sex chromosome instead of the XX found in normal females. The effects of this syndrome include retarded growth and sexual development and physical deformities.

The remaining 70 percent of the cases, or 175,000, are infants affected by environmental toxins, which the mother had either ingested or taken in through the

environment as pollutants. One birth in 750 is estimated to be a fetal alcohol syndrome (FAS) infant. Caused by maternal alcohol abuse, FAS is the most common cause of mental retardation; other effects include altered facial features and hyperactivity.

HIGH-RISK PREGNANCIES

Some women are more likely to give birth to an infant with the capacity to develop physical, intellectual, emotional, or social handicaps. Such handicaps may be due to unfavorable heredity, environmental influences, or both. The age of the mother at either end of the reproductive continuum is one factor. Teenagers, especially those under 16 when conception occurs, and women who are 40 years of age or older when conception occurs place their pregnancies at risk. For example, women over the age of 40 are 11.5 times more likely to have a Down syndrome child than women who are 29 years old or younger. Other risk factors include low socioeconomic status (which may be an indicator of poor maternal and/or prenatal care), nutrition, maternal diabetes or hypertension, birth of a previous infant with anomalies, and/or maternal drug use during pregnancy.

Continuum of Reproductive Deficit. This concept was first used by Pasamanick and Lilienfeld (1955) to describe the effect on the fetus from most to least severe: death; malformation; growth retardation; and functional deficits, including mental retardation. Which of these will occur in any individual case depends on many factors. For example, the genotype (inherited component) of the mother may act as a safeguard: Only 20 percent of the infants whose mothers took the prescribed tranquilizer Thalidomide during pregnancy in the 1950s were born with the extreme physical defects (missing or incomplete limbs) associated with that sedative. Twenty percent of pregnancies end in spontaneous abortion or miscarriage during early pregnancy; most of these are malformed embryos, with 25 to 50 percent of them having chromosomal abnormalities incompatible with life.

Principles of Teratology. According to J. G. Wilson (1973), who first described the role of environment in birth defects, the genotype, the develop-

mental stage of exposure, the dose of the agent, and the pattern of exposure all determine the effect on the fetus. The most sensitive time is during major organogenesis (embryogenesis), in the first trimester (three months of fetal development), when the major organs are developing. Growth retardation and mental retardation are the common results of late fetal exposure, during the last trimester. Death and major malformations occur with exposure during the first trimester, but, since the central nervous system develops during the last third of pregnancy, this is when cognitive deficits are most likely to occur. The critical-periods, or critical-moments, hypothesis (Stockard, 1921) states that the part of the brain that is developing at the time of exposure is most affected. Other parts may be affected as well, but to a lesser extent. Some agents may have an effect on a specific organ system, whereas another agent will not. For example, the heart develops early, with the first heartbeats at 22 to 28 days of age. Cardiac anomalies are one of the consequences of maternal alcohol, cocaine, and/or amphetamine use. The amount of the agent to which the fetus is exposed and the pattern of maternal ingestion are both important. The pattern of binge drinking, for example, is probably more important for the development of FAS than is the total amount of alcohol consumed during the pregnancy. This is particularly true if the binges occur during the first few weeks of gestation, usually before a woman is aware that she is pregnant. Nicotine (cigarettes, snuff) does not have a detrimental effect upon fetal growth until the second trimester, that is, beginning with the fourth month. All forms of tobacco ingestion result in growth-retarded infants. There is evidence that subsequent learning rates are affected as well (Martin, 1992).

Causal Factors. Genetic factors and chromosomal abnormalities have been mentioned as part of the inherited constitution of the mother and father. Although usually they can be assessed prior to delivery of the infant, they cannot be altered. Chronic illness of the mother is another factor. Maternal diabetes, epilepsy, and heart disease may have an effect on the infant, since prescription medications usually must be taken by the pregnant woman to control these conditions. These drugs then pass through the placenta to enter fetal blood through the umbilicus.

Low-birth-weight infants are at considerable risk, particularly if the infant is also premature. Low birth weight is defined as under 2,500 grams, or 5.5 pounds. An infant below a birth weight of 2,000 grams, or under 4.5 pounds, is at risk for neonatal (newborn) death.

A number of prescription drugs place the fetus at risk; thus it is important for doctors to ascertain whether a woman is—or might be—pregnant before prescribing them. Over-the-counter drugs (e.g., aspirin) and drugs of abuse (e.g., alcohol) are causal agents over which the pregnant woman has control. Other environmental agents, such as pollutants and food additives, may be only partially avoidable. Awareness of such factors and care to avoid them are both necessary.

Maternal Role. Maternal drug ingestion does not affect the infant unless the placental barrier is crossed. The placenta is the organ that supplies the fetus with all nutrients. One scientist has said that the placental “barrier” should be relabeled “sieve,” since all molecules (except the very large ones, such as protein molecules) do cross. The rate of flow increases with placental and fetal age, so the older fetus is exposed to more of a drug. Flow rate is important, because drug uptake by the fetus occurs only if the placenta is unable to metabolize and detoxify the drug. This rate depends on the drug itself, the age of the placenta (and fetus), and the physical condition of the mother. Some of the drugs that cross the placenta easily and that have demonstrated adverse physiological or chemical effects on the fetus include nicotine (cigarettes, snuff), alcohol, caffeine (coffee, tea, cola drinks), amphetamines (particularly methamphetamine), LSD, phenobarbital, morphine, heroin, and antihistamines. Some of the prescription drugs that cross readily include androgens, estrogens, corticosteroids, reserpine, cortisone, streptomycin, sulfonamide, Thalidomide, vitamin K (in excess), and some diuretics. Some fertility drugs (sex hormones), notably DES, may result in genital malformations in male and female fetuses during weeks four to twelve of pregnancy and vaginal cancer in female offspring if taken up to week eighteen in pregnancy. Although women stop taking fertility drugs once they become pregnant, they may not be immediately aware they are pregnant. The ef-

fects of exposure of the pregnant women to heavy metals on later functioning and survival of the offspring have ranged from extreme physical and mental disability in the case of methylmercury exposure (the Minamata syndrome) to possible mental retardation from low levels of lead. Physical agents, such as X rays, tend to be partially screened out by maternal tissue, but chemical agents are transported across the placenta without such a screening effect.

The mechanism of individual drug action is not always known. It is unclear if the effects of cocaine on infant small size, small head circumference, and hyper- or hypoactivity are due to the direct effects of the drug on cellular growth and differentiation or if cocaine indirectly affects vascularization by lessening placental blood flow. The latter is true of nicotine, which, like cocaine, activates the central nervous system (CNS). Alcohol affects cellular growth very early in pregnancy, and death of the fetus and/or severe brain malformations result when intake is high enough.

Paternal Role. Very little research has been done in the area of paternal effects. Although drugs taken by the mother can affect the fetus during the entire gestational period, drugs taken by the father can have an effect only by affecting the sperm prior to conception and the germ cells at the time of conception itself. The only way that the fetus can be affected is by changes in the sperm. These changes may affect sperm morphology (genes, chromosomes, cell physiology), count (fewer viable sperm), or motility (ability to move up the vaginal tract). All of these have been found with heavy alcohol, nicotine, cocaine, and heroin use. Older paternal age and poor physical condition have been correlated with an increase in the incidence of Down syndrome in infants sired by German World War II soldiers who were survivors of Allied prison camps. Smoking reduces sperm count and motility, and one large-scale but unreplicated study found increases in facial malformations in the offspring of smoking fathers. The effects of other drugs on offspring function are not known at this time. Methadone is excreted in high concentrations in human semen, however, and caffeine, lead, and anesthetic gases are other agents that may affect human offspring. Studies on these agents need to be performed (Soyka & Jaffe, 1980).

COGNITIVE AND BEHAVIORAL DEFICITS

Several scientists have argued that mental and other functional effects on the infant and child depend not upon the specific drug, but upon the amount of the agent and the stage of gestation when it was delivered. For example, in utero exposures to amphetamines, nicotine, and alcohol all result in a higher rate of activity in the offspring. Since amphetamines and nicotine are CNS-activating agents and alcohol is a sedative, this effect would not be understandable if one looked only at pharmacology. One of the most potent predictors of problems that emerge in childhood is an infant's being growth-retarded. An infant that fails to thrive usually has developmental slowing, which in turn results in slower learning and delayed developmental milestones. Hundreds of studies have confirmed that the mother's smoking during the last two trimesters is the strongest predictor of small size at birth. There is a dose-response effect—that is, size decreases as cigarette intake increases, with the heaviest smokers delivering the most growth-retarded infants. Cocaine, alcohol, amphetamine, and heroin exposure also result in undergrown newborns. One test frequently given to newborns is the Brazelton neonatal scale, which measures reflexes, babies' abilities to console themselves, and responses to auditory and visual cues.

One indication of delayed neurological maturation is the failure to habituate to (get used to) a stimulus, which indicates a lack of learning. Maternal alcohol use has this effect on the baby. Activity is a measure of arousal and rest patterns. Amphetamine exposure results in overactive infants, and animal studies have found that this hyperactivity continues until old age. Such lifespan (longitudinal) studies have not been performed on humans. One clinical study found that clinically diagnosed hyperactive children were more likely to have had mothers who smoked cigarettes during their pregnancies.

Nicotine and amphetamines are similar drugs. Cocaine exposure results in either very high or very low activity patterns, which are not normal and which interfere with learning. The male fetus is the one most at risk for developmental disabilities, and hyperactivity

is found in boys far more often than in girls. Learning and performance in offspring exposed in utero have been shown to be affected by smoking, alcohol, barbiturates, and amphetamines. One long-term British study has found lowered scores on achievement tests in children whose mothers smoked during pregnancy. This was particularly true of mathematics scores through age 11, although the effect lessened with age. This study was flawed, however, because questions were not asked about use of other drugs, such as alcohol. The effects could have been due to an interactive effect of multiple drug use (Martin, 1987).

REVERSIBILITY OF DEFICITS

The human brain continues to grow and develop after birth and has a great capacity for overcoming deficits. The only long-term studies done on humans are studies of moderate maternal alcohol use; these continue to find vigilance, attention, and memory deficits up through age 11 (Carmichael-Olson et al., 1992). A Swedish study on maternal amphetamine users, the only long-term study done on this drug, found lower IQ scores at age 4 in children whose mothers used drugs during pregnancy as compared with children of mothers who did not use drugs. Many animal studies have found irreversible deficits in offspring learning, attention, ability to withstand stress, and earlier death for the offspring of dams who were administered moderately high to high levels of drugs.

The long-term effects on offspring of maternal drug and other substance use are unpredictable, since so many other factors may enter into the equation. There is no reason to use drugs during pregnancy and many reasons for not using them. Certain prescription medications are the exceptions, since to discontinue them is to place the health of the mother at risk. In such cases, the doctor has an ethical obligation to fully inform the mother of possible effects on the fetus.

BIBLIOGRAPHY

- CARMICHAEL-OLSON, H., SAMPSON, P. D., BARR, H. STREISSGUTH, A. P., & BOOKSTEIN, F. L. (1992). Prenatal exposure to alcohol and school problems in late childhood: A longitudinal prospective study. *Development and Psychopathology*, 4, 341-359.

- MARTIN, J. C. (1992). Maternal tobacco product- or amphetamine-use and offspring effect. In T. B. Sondregger (Ed.), *Perinatal substance abuse: Research findings and clinical implications*. Baltimore: Johns Hopkins Press.
- MARTIN, J. C. (1984). Perinatal psychoactive drug use: Effects on gender, development, and function in offspring. In T. B. Sondregger (Ed.), *Psychology and Gender* (pp. 227–266). 32nd Annual Nebraska Symposium on Motivation. Lincoln: University of Nebraska Press.
- MARTIN, J. C. (1987). Multiple prenatal drug exposure and fetal outcome. In T. Fujii & P. Adams (Eds.), *Functional teratogenesis*. Tokyo: Teikyo University Press.
- MILLEN, J. W. (1963). Timing of human congenital malformations: With a timetable of human development. *Developmental Medicine and Child Neurology*, 5, 343–350.
- PASAMANICK, B., & LILIENFELD, A. M. (1955). Association of maternal and fetal factors with the development of mental deficiency: I. Abnormalities in the prenatal and perinatal periods. *Journal of the American Medical Association*, 159, 155–160.
- SOYKA, L. E., & JAFFE, J. M. (1980). Male-mediated drug effects on offspring. In R. Schwartz & S. J. Yaffee (Eds.), *Progress in clinical and biological research: Drug and chemical risks to the fetus and newborn* (pp. 49–66). New York: A. Liss.
- STOCKARD, C. R. (1921). Developmental rate and structural expression: An experimental study of twins, “Double monsters,” and single deformities, and the interaction among embryonic organs during their origin and development. *American Journal of Anatomy*, 28, 115–227.
- WILSON, J. G. (1973). *Environment and birth defects*. New York: Academic Press.

JOAN C. MARTIN

BIRTH ORDER, SPACING, AND FAMILY SIZE Do big families necessarily make dumb children? Are later-born children generally at an intellectual disadvantage compared to their earlier born siblings? Does wider spacing between children result in smarter children? Contrary to a great deal of popular belief (and even some expert opinions), the answer to each of these three questions is a fairly unequivocal no. This article will treat the empirical and theoretical literature that links family structure and intelligence.

In 1874, Sir Francis GALTON published *English Men of Science*. Galton’s book noted that most British scientists were the firstborn in their families. His observation anticipated an ever-increasing interest in the relationship between structural features of the family and intelligence. (Strictly speaking, Galton was interested in scientific eminence and not intelligence. The two are certainly correlated, however, and in this article we will treat several related areas, including intelligence, achievement, and eminence.) That interest continues unabated, over one hundred years later.

DEFINITION AND CLASSIFICATION

Birth order, family size, and other measures of family structure have been tremendously popular research variables. Rodgers and Thompson (1985, p. 144) defined family structure characteristics as ones “related to the age, sex, and presence or absence of human members of the family (or household) environment.” Although birth order and family size have been the most popular family-structure variables—especially for those studying intelligence—hundreds of other measures fall within the scope of this definition. Other family-structure variables used by family researchers include parental age at first birth, inverse family size, average family size, average time between births, family density, family sex ratio, paternal presence in the household, number of adults in the household, and sex of the next older sibling. Most of these measures were constructed by researchers who recognized that birth order and family size are not always very sensitive measures of the types of psychological and sociological processes that go on within families.

Rodgers and Thompson (1985, p. 150) defined two different ways to classify such measures. The first distinguished between static and dynamic measures. Static measures (e.g., birth order, spacing to the next older sibling, mother’s age at first birth) are ones that, once defined, stay fixed. Dynamic measures change over time with the changing family (e.g., family size, number of younger siblings, father’s age at last birth). The second classification distinguished between measures of individuals and measures of families. Birth order is the prototypical measure of an individual; such indices can be used to distinguish between siblings within a family. Family size is an example of a family

measure; such measures are useful for studying processes that make families different from one another.

FAMILY-STRUCTURE VARIABLES— CAUSES OR CONFOUNDERS?

In fact, much of the confusion over what causal influences family-structure variables have on intelligence (and other dependent variables) can be traced to confusion over family and individual characteristics. Most family-structure variables—even very fine-tuned individual-level measures—can act as proxies for higher-level sociological processes. The literature relating family structure to intelligence is filled with misattributed causal interpretations in which processes occurring at the between-family level are mistakenly believed to contribute to differences between siblings. This point is important enough to justify an example.

Suppose that parents with low levels of formal education tend to have larger families than do parents with higher levels of formal education. (Note that there is considerable empirical evidence supporting this particular pattern. However, low-educated and low-socioeconomic status individuals also are less likely to marry and have any children. Higgins, Reed, and Reed [1962] used this latter relationship to help explain why aggregate intelligence is not going down.) This would imply that children in families of ten members, for example, would come disproportionately from low-educated parents. To the extent that education and achievement scores correlate, one might conclude that large families *cause* lower achievement, when, in fact, an equally plausible interpretation is that low education *causes* both large families and low intelligence. By the same argument, in a cross-section of children, those with birth orders of five or higher would come disproportionately from low-educated parents; in this case birth order would act as a proxy for parental education, just as family size did in the previous example. Only recently in the long history of family-structure research has the importance of longitudinal and intact-family data been fully appreciated. Such data are required before the confounded relationships between family structure variables and other more sociological variables (e.g., socioeconomic status, parental education, parental income) can be understood.

METHODS OF STUDYING FAMILY-STRUCTURE EFFECTS

There are three ways to study birth order and other family-structure variables. The first two are empirical approaches. The third combines theory development with empirical testing of a theory. In the first method, a researcher selects a special group of interest, and collects data to investigate whether the birth-order patterns in that special group are the same as those in the population as a whole. Birth-order studies have been conducted of alcoholics, artists, assassins, authors, convicts, dentists, hockey players, musicians, presidents, scientists, smokers, soldiers, and strippers, among many others. Even the birth-order patterns among birth-order researchers have been studied. Galton's work was an early example of this approach to studying the relation between family structure and intelligence (as measured by scientific eminence). The second method is to define some quantitative dependent variable (like intelligence) and test analytic models (e.g., correlation or regression) relating measures of family structure to the dependent variable.

The third method of research, and the most scientific, involves the development of formal models (social-science theories) that explain the relationship between family structure and some important dependent variable like intelligence; then these models are tested against patterns in empirical data. Such formal models were not proposed in the family-structure literature until the 1970s. Several have engendered great interest and controversy among social scientists; these will be reviewed later in this article.

EMPIRICAL STUDIES

A number of reviews have considered the empirical literature relating family structure to intelligence (e.g., Adams, 1972; Altus, 1966; Anastasi, 1956; Rodgers & Thompson, 1985; Schachter, 1963; Schooler, 1972; Thompson, 1974; Zajonc, Markus, & Markus, 1979). Two recent books have addressed these issues (Blake, 1989; Ernst & Angst, 1983). Many reviews are at least slightly skeptical that there is a strong relationship between family structure and intelligence: "... the general lack of consistent [birth order] findings ... leaves real doubt as to whether the chance of positive results

is worth the heavy investment needed to carry out any more definitive studies” (Schooler, 1972, p. 174); “. . . findings suggest that it is not size of sibship *per se* but other factors associated with family size within a given culture which produce the obtained differentials in intellectual level” (Anastasi, 1956, p. 197).

Perhaps the best known of the empirical studies relating family structure and intelligence was conducted by Belmont and Marolla (1973). They obtained data from approximately 400,000 19-year-old Dutch males who were in the military and who had been born between 1944–1947. For each respondent Belmont and Marolla had both family-structure information (birth order and family size) and an IQ measure (from the Raven Progressive Matrices). A portrayal of their famous plot of the birth order/family size/IQ pattern is shown in Figure 1 (note that low Raven scores indicate high IQ). Clearly, there is a tendency for intelligence to decline with increasing birth order and family size. When the respondents were separated into three social classes, the patterns persisted.

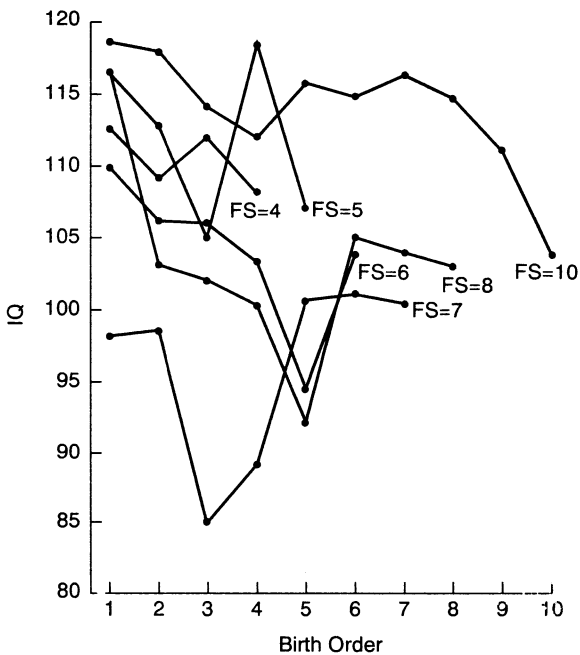


Figure 1
IQ by birth order and family size (FS) among 19-year-old Dutch males
SOURCE: Belmont & Marolla, 1973.

A number of other national studies showed similar patterns; several of these are reviewed and shown graphically in Zajonc (1976). Because these data sets are so large, many believed that methodological artifacts present in previous smaller and more select data sets could simply not be present in patterns such as those from Belmont and Marolla’s data. Zajonc (1975, p. 37), writing in a popular magazine, reviewed the patterns in the Dutch data and suggested to parents that “to have brainier children, keep them few and far between. The brightest children come from the smallest families, and within any given family size, the children that come along early tend to have higher IQs.” Unfortunately, when these words were written few intact-family studies (studies in which children *within the same family* were compared) had been conducted. When such data were examined (e.g., Berbaum &

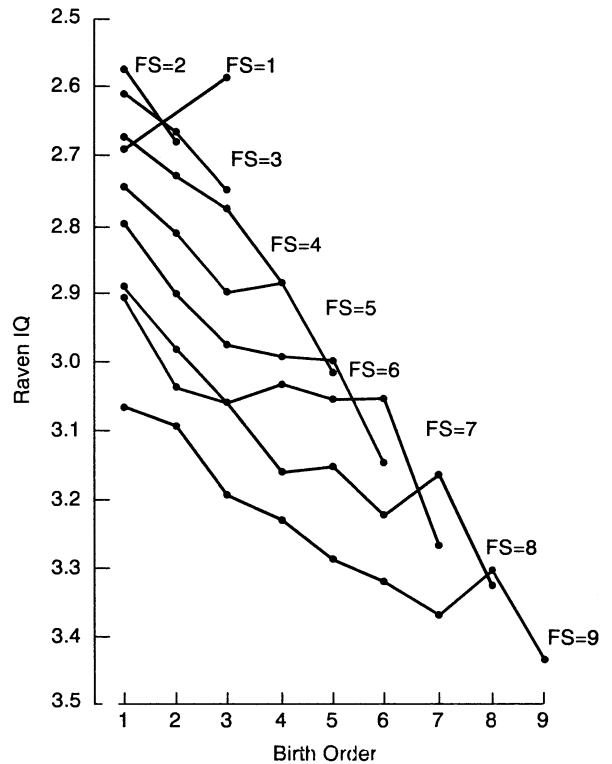


Figure 2
IQ by birth order and family size (FS) among U. S. children from intact families
SOURCE: Outhit, 1933, cited in Berbaum & Moreland, 1980.

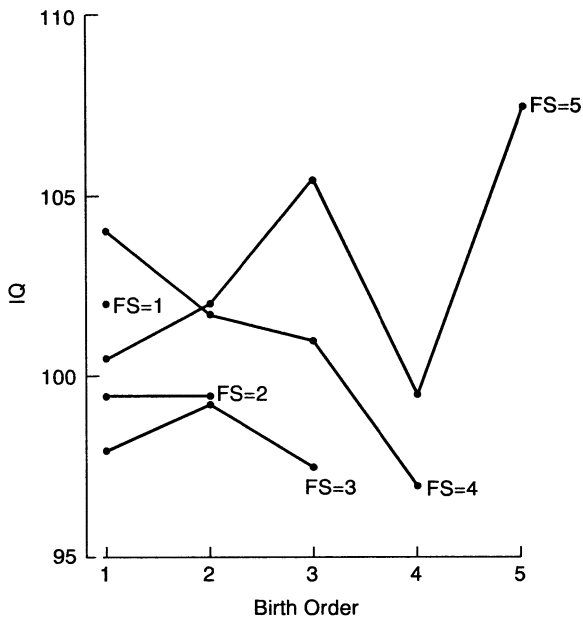


Figure 3

IQ by birth order and family size (FS) among Ohio children from intact families in Fels data

SOURCE: Rodgers, 1984.

Moreland, 1980, p. 509; Rodgers, 1984, p. 327), the patterns were totally different from those systematic relationships exemplified by the Dutch data (see Figures 2 and 3). In these families, which were relatively homogeneous in terms of SES and parental education, neither birth order nor family size had any strong relationship to IQ scores.

One of the peculiar features of the Belmont and Marolla data was the large drop-off in the IQ scores of last-born children (which can be seen at the end of each line in Figure 1). Zajonc (1976) suggested that this was caused by a "tutoring effect," that gave all but the last sibling the opportunity to teach a younger child, but last-born children do not have a younger sibling to teach. Blake (1981), on the other hand, argued that the last-born discontinuity was probably caused by a potato famine that occurred during the birth years of the Dutch sample, which disproportionately affected the poor and lower-class; those who died or suffered malnutrition stopped having children, so that last-borns were more likely to come from families strongly affected by the famine. These two different

arguments illustrate how plausible explanatory models can be developed at a number of levels, from the within-family theorizing of Zajonc to the sociological explanation offered by Blake. Intact-family and longitudinal data are required to separate the different predictions of such competing explanatory models.

THEORIES OF THE RELATIONSHIP BETWEEN FAMILY STRUCTURE AND INTELLIGENCE

It is only in the last two decades that formal theories have been developed. The first such theory, the confluence model (Zajonc & Markus, 1975), engendered a remarkable amount of discussion and controversy. Much of this controversy sprang from the same question that had muddied interpretation of the earlier empirical patterns: Are such patterns caused by within-family processes or by extra-family processes for which within-family measures are proxies?

The confluence model was developed directly from observation of the patterns in the Belmont and Marolla data and other similar national datasets. The model suggested that within any family there is an intellectual environment (the "confluence" of the individual contributions to it). An individual's contribution can be quantified by their mental age, which increases systematically with age. Thus, the presence of many children will dilute such an environment, and the developing children in a large family are at an intellectual disadvantage. Further, the theory suggested a tutoring effect that puts last-born children at an added disadvantage. Several theoretical features of the confluence model made it attractive. First, it was the first process-oriented theory to make quantitative predictions. Second, it accounted for the aggregate patterns in a number of previous datasets. Third, it accounted for spacing patterns as well as birth order and family size; the larger the spacing between children, the less the intellectual environment would be diluted by the presence of many young children. Finally, it also was consistent with other intelligence patterns in the literature (e.g., twins have lower IQs on average than nontwins).

Early acclaim for the confluence model was followed by criticism from a number of social scientists. Dozens of tests of the confluence model were per-

formed, including several using the type of intact-family data ideally necessary to test the model (e.g., Berbaum & Moreland, 1980; Galbraith, 1982; Grotevant, Scarr, & Weinberg, 1977; Rodgers, 1984; Retherford & Sewell, 1991). Many of these criticisms were concerned with the gap between the level at which the confluence model made predictions (i.e., it is a within-family model) and the patterns (between-family patterns) that had been observed to stimulate the development of the model: "Valid inferences regarding sibling differences in intellectual development cannot be reliably made from between-families samples" (Galbraith, 1982, p. 166). Steelman (1985, p. 353) made an extensive review of the merits and criticisms of the confluence model and concluded "studies reviewed in general tend to refute the confluence model."

However, without regard to the many subtle substantive and methodological arguments about the confluence model, Zajonc and Markus's (1975) work must be viewed as the signal development that stimulated theories of family structure and intelligence. Page and Granden (1979) advanced an "admixture hypothesis" suggesting that the apparent within-family patterns in national data sets are caused by mixtures of SES and educational backgrounds, and not by within-family processes at all. They noted that aggregate R^2 's of .60 to .90 dropped down to around .01 when the same data were analyzed at the individual level. Blake (1981) proposed a "dilution model," which suggested that the more children in an environment, the more diluted the parental attention and resources. While not as strongly quantitative as the confluence model, her explanation also moved to the within-family level to explain aggregate patterns in intelligence (which patterns disappeared in the subsequent studies using intact-family data). Rodgers and Rowe (1985) proposed a "contiguity model," which suggested that children who shared a great deal of the family environment (such sharing was measured by birth order, age spacing, and sex similarity) should be more similar in intelligence. When they tested this model using within-family data, however, no consistent patterns emerged.

When the proper analyses are run, the strong relationship between these family-structure variables and intelligence that can be observed in aggregate be-

tween-family data virtually disappears in within-family data. This observation alone helps us to understand causal mechanisms that underlie the consistent patterns in national data like those from Belmont and Marolla's (1973) work. Within-family processes (like those suggested by the confluence model or the dilution model) are not potent predictors of intellectual differences within families. Between-family differences in aggregate data account for most of the fit of the confluence and dilution models.

CONCLUSION

There is little evidence that structural features of the family, such as birth order, family size, and spacing, have direct causal influences on the intelligence of children. If such influences exist, they are certainly subtle and will require new developments in theory and measurement. A rather more fruitful research enterprise than continued focus on strictly structural variables would be to develop models that are thoughtful about the mechanisms of influence on intelligence. Environmental influences are probably translated to children through much more process-oriented mechanisms than are measured by such structural variables; for example, it seems very reasonable that the amount of time a mother reads to her child should be a more potent influence on the child's intelligence and achievement than the child's birth order per se.

However, the family structure measures carry with them hidden information about parental education, socioeconomic status, intrauterine effects, and other psychological, sociological, and biological processes. Understanding these correlates (and artifacts) can substantially enrich our understanding of the different factors that directly and indirectly affect intelligence.

Given the general conclusion of this article—that most past fascination with the influence of family structure on intelligence has been driven by artifacts—we might wonder why both researchers and informed laymen alike have such a fascination with the relation between family structure and intelligence. We speculate briefly about this fascination. Family structure features are salient, often discussed, and superficially significant. Everyone has a birth order, and it is

easy to observe and talk about. Out of such discussions arise hundreds of informal, personal theories of the birth order–intelligence relationship. Other influences that certainly contribute causally to intelligence and achievement (factors like genetic characteristics, quality of schooling, quality of parental support) are more subtle, less obvious, and more difficult to discuss.

Nevertheless, more than a century of research on family structure is enough to demonstrate that the most potent sources of influence on intelligence are not those measured structurally. We need to look well beyond the too-simple horizon defined by birth order, family size, and spacing to understand the causal influences that contribute to the intellectual development of our children.

(See also: FAMILY ENVIRONMENTS; NATURE, NURTURE, AND DEVELOPMENT.)

BIBLIOGRAPHY

- ADAMS, B. (1972). Birth order: A critical review. *Sociometry*, 35, 411–439.
- ALTUS, W. D. (1966). Birth order and its sequelae. *Science*, 151, 44–49.
- ANASTASI, A. (1956). Intelligence and family size. *Psychological Bulletin*, 53, 187–209.
- BELMONT, L., & MAROLLA, F. A. (1973). Birth order, family size, and intelligence. *Science*, 182, 1096–1101.
- BERBAUM, M. L., & MORELAND, R. L. (1980). Intellectual development within the family: A new application of the confluence model. *Developmental Psychology*, 16, 506–515.
- BLAKE, J. (1981). Family size and the quality of children. *Demography*, 18, 421–442.
- BLAKE, J. (1989). *Family size and achievement*. Berkeley: University of California Press.
- ERNST, C., & ANGST, J. (1983). *Birth order: Its influence on personality*. New York: Springer Verlag.
- GALBRAITH, R. C. (1982). Sibling spacing and intellectual development: A closer look at the confluence model. *Developmental Psychology*, 18, 151–173.
- GALTON, F. (1874). *English men of science*. London: Macmillan.
- GROTEVANT, H. D., SCARR, S., & WEINBERG, R. A. (1977). Intellectual development with adopted and natural children: A test of the Zajonc and Markus model. *Child Development*, 48, 1699–1703.
- HIGGINS, J. V., REED, E. W., & REED, S. C. (1962). Intelligence and family size: A paradox resolved. *Eugenics Quarterly*, 9, 84–90.
- OUTHIX, M. C. (1933). A study of the resemblance of parents and children in general intelligence. *Archives of Psychology*, 149, 1–60.
- PAGE, E. B., & GRANDEN, G. (1979). Family configuration and mental ability: Two theories contrasted with U.S. data. *American Educational Research Journal*, 16, 257–272.
- RETFERFORD, R. D., & SEWELL, W. H. (1991). Birth order and intelligence: Further tests of the confluence model. *American Sociological Review*, 56, 141–158.
- RODGERS, J. L. (1984). Confluence effects: Not here, not now! *Developmental Psychology*, 20, 321–331.
- RODGERS, J. L., & ROWE, D. C. (1985). Does contiguity breed similarity? A within-family analysis of nonshared sources of IQ differences between siblings. *Developmental Psychology*, 21, 743–746.
- RODGERS, J. L., & THOMPSON, V. D. (1985). Toward a general framework of family structure: A review of theory-based empirical research. *Population and Environment*, 8, 143–172.
- SCHACHTER, S. (1963). Birth order, eminence, and higher education. *American Sociological Review*, 28, 757–768.
- SCHOOLER, C. (1972). Birth order effects: Not here, not now! *Psychological Bulletin*, 78, 161–175.
- STEELMAN, L. C. (1985). A tale of two variables: A review of the intellectual consequences of sibship size and birth order. *Review of Educational Research*, 55, 353–386.
- THOMPSON, V. D. (1974). Family size: implicit policies and assumed psychological outcomes. *Journal of Social Issues*, 30, 93–124.
- ZAJONC, R. B. (1975, January). Dumber by the dozen. *Psychology Today*.
- ZAJONC, R. B. (1976). Family configuration and intelligence. *Science*, 192, 227–236.
- ZAJONC, R. B., & MARKUS, G. B. (1975). Birth order and intellectual development. *Psychological Review*, 82, 74–88.
- ZAJONC, R. B., MARKUS, H., & MARKUS, G. B. (1979). The birth order puzzle. *Journal of Personality and Social Psychology*, 37, 1325–1341.

JOSEPH LEE RODGERS
DAVID C. ROWE

BLOCK DESIGN See WAIS-R SUBTESTS.

BOND SAMPLING THEORY OF HUMAN ABILITIES Bond sampling theory was developed by Godfrey H. THOMSON to explain the g factor that Charles SPEARMAN specified for ability test scores (see Spearman's TWO-FACTOR THEORY). Spearman showed that if all of many tests involved one ability in common, which he symbolized as g , and each test involved also a specific ability that was not involved in any other test, then the matrix of intercorrelations among such tests would stand in a very precise order that he designated "hierarchical order" (not to be confused with HIERARCHICAL THEORIES OF INTELLIGENCE). In particular, under these conditions, if the tests in a matrix of intercorrelations are arranged in order of the sum of their correlations with all other tests in the battery (from largest sum to smallest sum), it can be seen that the correlations in any two columns have a constant ratio throughout all rows. For example, if the tests in rows i and j and columns k and m are arranged

$$\begin{array}{cc} r_{ik} & r_{im} \\ r_{jk} & r_{jm} \end{array},$$

then the ratio r_{ik}/r_{im} is equal to the ratio r_{jk}/r_{jm} . This is true for all tests ($i = 1, 2, \dots, n$) throughout the matrix of test intercorrelations. The hierarchical order is the order from the test involving the largest amount of the common factor (the sum of its correlations with all other tests being the largest) to the test with smallest amount of the common factor (its sum of correlations being the smallest).

Since the ratios r_{ik}/r_{im} and r_{jk}/r_{jm} are equal, it follows that the following difference, which Spearman called the "tetrad difference," is zero:

$$(r_{ik})(r_{jm}) - (r_{im})(r_{jk}) = 0. \quad (1)$$

A hypothesis that tests do indeed involve the two kinds of factors stipulated in the model is supported if all the tetrad differences are found to be zero (within chance variability). Since for n tests there are $n(n-1)(n-2)(n-3)/8$ of these differences, and all must be zero, the test is very demanding. For example, for 10 tests there are $(10)(9)(8)(7)/8 = 630$ tetrad differences, all of which must be zero.

Under conditions in which the model hypothesis is supported, Spearman interpreted g as "mental energy" and specific abilities as "neural machines" that made use of this energy. In a series of articles beginning in 1916, Thomson challenged this interpretation. He argued that hierarchical order among test intercorrelations is a mathematical consequence of tests overlapping in their sampling of many thousands of components of the mind, which he called "bonds." Thomson argued that "factorists" such as Spearman were reifying a mathematical concept—that is, ascribing reality to what was merely a statistical coefficient.

THOMSON'S ALTERNATIVE EXPLANATION

Both Thomson and Spearman aimed to explain a mathematical (statistical) phenomenon, hierarchical order. But in Thomson's interpretation of this phenomenon, it was proposed that the expected value of the correlation between any two tests a and b is

$$r_{ab} = \sqrt{p_a \cdot p_b} \quad (2)$$

where p_a and p_b are the proportions of the mind's bonds that are sampled by tests a and b , respectively. The expected value of the tetrad difference for a set of tests fitting this model is zero (Thomson, 1927), just as in Spearman's theory. Thus, Thomson argued that hierarchical order was not the consequence of any psychological law but rather the consequence of the laws of probability.

Although he did not commit himself to a particular interpretation of these bonds, Thomson suggested a neurophysiological model such that bonds represented connections among neurons and that cognition represented patterns of excitation of these neurons. Intelligence, he offered, was "probably associated with the number and complexity of the patterns which the brain can (or could) make" (Thomson, 1939, p. 51), a hypothesis similar to one by E. L. THORNDIKE (1925).

Thomson argued against faculty theories in which the mind was held to be highly structured and composed of a few factors that operate through an immense number of specific neural machines. Thomson believed that the tendency to hierarchical order and the presence of g suggested that the mind was largely unstructured, in the sense of having no fixed or strong

linkages between its components, so that any sample of components could be established to perform cognitive tasks such as intelligence tests. Deviations from this order suggested structure in the form of group factors (common factors underlying three or more tests). Thomson proposed that structure of this type resulted from development and education, noting that adults' test scores required more factors to explain than children's test scores (cf. Atkin et al., 1977).

Thomson (1952) contended that the number of bonds a person possessed had multiple determinants, notably genes, informal learning, schooling, and other environmental influences. Given this myriad of influences, the intercorrelation among test scores would suggest performance depended on a small number of common factors (plus specifics). Yet, it is just as likely that commonalities among tests are due to overlapping samples of bonds. This does not prove that common factors do not exist, but it does suggest that they are "fictions" or hypothetical entities. Thomson allowed the possibility that g and other common factors would nonetheless be useful for some practical purposes.

SIMILAR THEORETICAL PERSPECTIVES

Theories similar to Thomson's were put forth by R. C. Tryon (1935), G. A. Ferguson (1954), L. G. HUMPHREYS (1984), and A. E. Maxwell (1972). Humphreys (1971, p. 31), one of Thomson's principal proponents, defined intelligence as "the entire repertoire of acquired skills, knowledge, learning sets, and generalization tendencies considered intellectual in nature that are available at any one period in time." Humphreys believed circularity was avoided in this definition because there was sufficient agreement among psychologists as to what types of test items were appropriate. Humphreys's definition suggests that intelligence tests should be designed to sample as much of these acquired components as possible and that an omnibus test consisting of a wide variety of item types would be a better measure of intelligence than a test consisting of one item type focusing on one theorist's notion of intelligence.

Ferguson (1954) offered a learning-theory account of ability formation that is consistent with sampling theory. He proposed that positive correlations among

ability tests were due to transfer of learning, which occurred because tasks shared components. All abilities are "overlearned acquisitions," or skills that have reached their asymptote and have stabilized. When a person must learn a new task, task components that were acquired previously transfer to the new situation. With practice, task performance stabilizes and becomes an ability. Abilities that share a number of components form a common factor.

EVALUATION OF BOND SAMPLING THEORY

Bond sampling theory is primarily a mathematical account of general ability and other common factors. As a mathematical theory, it is a sound alternative to the two-factor theory and its successors (Maxwell, 1972). Thomson's point that factors are not immutable structures in the brain is certainly consistent with the proliferation of factors that has occurred in the decades since Spearman. If factors were robust phenomena and truly reflected underlying brain structures, then perhaps sample variables such as age, education level, sex, and ethnicity would not have such strong effects on factor solutions.

Bond sampling theory can be embarrassed by behavioral data. For example, the theory would predict that an elementary cognitive variable such as inspection time would not show much overlap with g , yet the median correlation with general ability measures was found to be $-.50$ in one meta-analysis (Kranzler & Jensen, 1989). Either the inspection-time task is not as elementary as it would appear to be and actually evokes a large number of bonds, or it serves as a major component of intelligence, and bond sampling theory is thus questioned.

Thomson believed that, in time, bond sampling theory would be demonstrated to be consistent with neurophysiology. The difficulty in evaluating this claim is that it is ill specified; one can only infer what Thomson's hypotheses were. There appear to be three major propositions. One is that the brain does not have fixed or strong linkages between its components, so that any sample of components could be established to perform the activity required by a test. The second is that brain structures do not exist that correspond to cognitive factors. The third is that intelligence, which largely

consists of acquired knowledge and skills, reflects the number and complexity of the patterns of excitation possible in the brain.

Knowledge of how cognitive processes map onto the brain is still rather primitive, the principal data coming from experimental work with animals and clinical studies of human brain injury. Nonetheless, the available evidence does suggest that the first proposition is true. The brain's modifiable synapses are relatively free to participate in learning, memory, and cognition (Brindley, 1969; John, 1972; Marr, 1970).

The second proposition is problematic with respect to the data. Specialized brain structures do exist, but they do not necessarily correspond to abilities found via factor analysis. For example, declarative memory (facts and episodes) and procedural memory (cognitive and motor skills) can be distinguished behaviorally and neurologically (Squire, 1987). Neuropsychological research has also demonstrated that cognitive functions can be normal even when certain memory abilities have been impaired. In cases of amnesia, memory of past events and school knowledge is intact, and immediate recall is intact if rehearsal is not interrupted. Learning new information is impossible if short-term memory capacity is exceeded. This phenomenon suggests specialized neural mechanisms for handling temporary and permanent memories.

The third proposition is not yet empirically testable. New brain imaging and monitoring techniques such as positron emission tomography, magnetic resonance imaging, and electroencephalograms will probably advance knowledge considerably. Even so, it seems unlikely that the means for estimating the number and complexity of neural connections ever will be discovered.

CONCLUSIONS

Bond sampling theory avoids some of the problems of essence and compound models (Horn, 1989) but has certain propositions that are untestable, given the present state of neuroscience. Perhaps the most important contributions of bond sampling theory are that it refutes faculty theories and discourages sole reliance on factor analysis in human abilities research, that it emphasizes the polygenetic and polyenvironmental determinants of abilities, and that it encourages co-

operation between psychology and neuroscience in understanding the complex phenomena of human cognition.

BIBLIOGRAPHY

- ATKIN, R., BRAY, R., DAVISON, M., HERZBERGER, S., HUMPHREYS, L. G., & SELZER, U. (1977). Ability factor differentiation, grades 5 through 11. *Applied Psychological Measurement, 1*, 65-76.
- BRINDLEY, G. S. (1969). Nerve net models of plausible size that perform many simple learning tasks. *Proceedings of the Royal Society (B)*, *174*, 173-191.
- FERGUSON, G. A. (1954). On learning and human ability. *Canadian Journal of Psychology, 8*, 95-112.
- HORN, J. (1989). Models of intelligence. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 29-73). Urbana: University of Illinois Press.
- HUMPHREYS, L. G. (1984). General intelligence. In C. R. Reynolds & R. T. Brown (Eds.), *Perspectives on bias in mental testing* (pp. 221-247). New York: Plenum.
- JOHN E. R. (1972). Switchboard versus statistical theories of human learning. *Science, 177*, 850-864.
- KRANZLER, J. H., & JENSEN, A. R. (1989). Inspection time and intelligence: A meta-analysis. *Intelligence, 13*, 329-347.
- MARR, D. (1970). A theory of cerebral neocortex. *Proceedings of the Royal Society (B)*, *176*, 161-234.
- MAXWELL, A. E. (1972). Factor analysis: Thomson's sampling theory recalled. *British Journal of Mathematical and Statistical Psychology, 25*, 1-21.
- SPEARMAN, C. (1927). *The abilities of man*. London: Macmillan.
- SQUIRE, L. R. (1987). *Memory and brain*. New York: Oxford University Press.
- THOMSON, G. H. (1927). The tetrad-difference criterion. *British Journal of Psychology, 17*, 321-336.
- THOMSON, G. H. (1939). *The factorial analysis of human ability*. Boston: Houghton-Mifflin. (Final version: 5th ed. London: University of London Press, 1951.)
- THOMSON, G. H. (1952). Godfrey Thomson. In E. G. Boring, H. S. Langfeld, H. Werner, & R. M. Yerkes (Eds.), *A history of psychology in autobiography* (Vol. 4, pp. 279-294). Worcester, MA: Clark University Press.
- THORNDIKE, E. L. (1925). *The measurement of intelligence*. New York: Bureau of Publications, Teachers College, Columbia University.
- TRYON, R. C. (1935). A theory of psychological compo-

nents: An alternative to mathematical factors. *Psychological Review*, 39, 324–351.

WILLIAM C. TIRRE

BRAIN Throughout the history of science, theories about human intelligence have been intricately entwined with theories about brain function. There have been questionable views, such as the notion that intelligence is related to the size of one's head. There have been genetic views, which suggest that intelligence is inherited by the transmission of "smart" genes. Finally, there have been environmental views, which suggest that intelligence is formed primarily by the sculpting of brain processes through experience. Clearly, both genetic and environmental factors contribute significantly to intellectual and brain development. But how does the brain's activity provide us with mental abilities, such as the ability to think, to understand, to remember, and to speak? How do different parts of the brain control these cognitive functions? How can an understanding of brain function help us understand human intelligence?

THEORIES OF INTELLIGENCE AND BRAIN FUNCTION

In the past 2.5 million years, the human brain has quadrupled in size, from the size of an ape's brain to that of modern humans. By evolutionary standards, this increase in brain size is quite rapid. During this time, we have also acquired skills that we conceptualize as "intellectual" (Calvin, 1990). In particular, we have acquired a sophisticated means by which to communicate with one another. These observations have led to the speculation that there is a relationship between brain size and intelligence. Enormous differences in brain size across species, as defined by the total mass of nerve cells in the brain, may be roughly correlated with differences in the complexity of brain processes among species. This theory also has been applied to the analysis of intelligence within a species, such as *homo sapiens*.

The view that differences in brain size in humans are related to differences in intelligence has a colorful and controversial history (Gould, 1981). In the nine-

teenth century, craniology, or the scientific study of head size, was used in anthropological studies to compare people of different races. It was suggested that people with larger heads, and, therefore, larger brains, were more intelligent. Indeed, in the nineteenth century a large head size was a popular means of proving one's intellectual prowess. Yet, this measure of intelligence was ultimately disregarded when some scientific studies failed to support socially accepted beliefs. For example, it was reported that the head size of a sample of hardened criminals was actually larger than average. Also, many prominent members of the intelligentsia actually had rather small heads. Apparently, head size was not a useful indicator of human intelligence.

Following the failure of head size as a biological measure of intelligence, some investigators turned to psychological measures of intelligence. One of the most prolific figures in the measurement of biological intelligence was Sir Francis GALTON, who was Charles Darwin's half-cousin. Galton studied psychological measures, such as reaction time, that is, the time it takes to respond to a stimulus (Galton, 1883). He believed that reaction time reflected the speed of neural activity and could, therefore, be a biological measure of the speed of intellectual thought. If true, Galton could then use the measure of reaction time to test his belief that intelligence is inherited (note here the influence of Darwin). The notion was appealing because intelligence could simply be related to a general brain property, such as neuronal speed or efficiency.

The notion that general neuronal efficiency plays a significant role in intellectual ability is still a viable theory. Charles SPEARMAN's theory (Spearman, 1923) characterized intelligence in terms of a general factor that he called the *g* factor. It was thought that intelligence was mediated primarily by this single, perhaps biological factor. It is reasonable to suppose that intellectual ability could be greatly influenced by the speed with which the tens of billions of nerve cells in the brain communicate with one another. Indeed, physiological measures, such as electroencephalographic (EEG) measures of nerve conductance velocities and positron emission tomographic (PET) measures of glucose metabolism, have been shown to be correlated with performance on IQ tests (Matarazzo, 1992).

A different perspective on the biology of intelligence can be viewed when one observes the breakdown of mental abilities due to brain injury or disease. In such neuropsychological investigations, the pattern of cognitive impairment after brain damage depends critically on the locus of brain damage. Indeed, the brain's architecture appears to be built in a modular or componential fashion (Kosslyn & Koenig, 1992). These modules or components include brain mechanisms for processes, such as object perception, spatial imagery, reading, language, movement, and memory. This concept of division of labor within the brain is based on the finding that brain damage in one location disrupts certain mental abilities but not others. For example, brain injury in one location can disrupt language but not spatial ability, whereas brain injury in another location can disrupt spatial ability but not language. Knowledge about the mental abilities associated with specific brain regions may provide an understanding of human intelligence. In particular, it may provide insight into the enormous diversity of intellectual abilities in humans. For example, some individuals are proficient in verbal skills, whereas others are proficient in spatial or mathematical skills. Still others are proficient in artistic skills, whereas others are proficient in musical skills. The next section briefly describes what is known about the anatomical basis of such intellectual skills.

THE ANATOMY OF MENTAL ABILITIES

The brain is not a uniform collection of nerve cells just as an automobile is not a uniform collection of metal. The brain has characteristic parts and connections that serve different functions (Kolb & Whishaw, 1990; Martin, 1989). Figure 1 shows a magnetic resonance image (MRI) of a living human brain. The MRI shows the medial or inner region of the brain imaged from the side. With this technique one can image both nerve cell bodies (i.e., gray matter) and nerve axons or fibers (i.e., white matter). Shown in Figure 1 is the cerebral cortex, the convoluted outer sheet of nerve cells that is vitally important for intellectual function. The cerebral cortex is divided into two hemispheres located on each side of the brain. Also shown are the white-matter axons of the corpus callosum, the fibers

that allow the two cerebral hemispheres to communicate with each other.

There are various ways by which the brain receives information from other parts of the body. For example, the brain stem, which is situated at the base of the brain (see Figure 1), connects the brain with the spinal cord and receives sensory information—such as pressure, temperature, and pain information—from other parts of the body. The brain stem is connected to subcortical structures, such as the hypothalamus, thalamus, and cerebellum, which serve to regulate bodily function (e.g., temperature, hunger, balance). Other sensory information is received from special receptors, such as the photoreceptors in the eye. Photoreceptors transmit nerve impulses to subcortical areas, such as the thalamus, which then relays the information to the cerebral cortex. In general, subcortical areas play a vital role in regulating bodily function and relaying information to be processed by the cerebral cortex. In many other animals, subcortical areas play a primary role in the range of behaviors exhibited by the animal. However, in humans, the cerebral cortex serves as the primary mediator of mental processes, such as sensation, thought, and action. It is this area of the brain that will be the major focus of this section.

The cerebral cortex is a sheet of nerve cells approximately 1.5 to 3.0 mm thick and 2,500 cm² in area (Kolb & Whishaw, 1990). If the surface of the cerebral cortex were flattened, it would be roughly equivalent to the area of a large pizza. It is the size of the cerebral cortex that has increased enormously in the last 2.5 million years. By contrast, the brain stem of a sheep or monkey is comparable in size to the brain stem of a human. The convoluted appearance of the cerebral cortex, with its ridges and valleys, is caused by squeezing a large surface area into the relatively small volume of the human skull. Underneath the cortical sheet of cell bodies is a massive and intricate array of axon fibers that connect regions within and between the two hemispheres. In general, each hemisphere operates or controls the opposite half of the body or visual space. That is, the left cerebral hemisphere controls muscles on the right side of the body. It also receives visual input presented to the right of eye fixation. Conversely, the right cerebral hemisphere controls

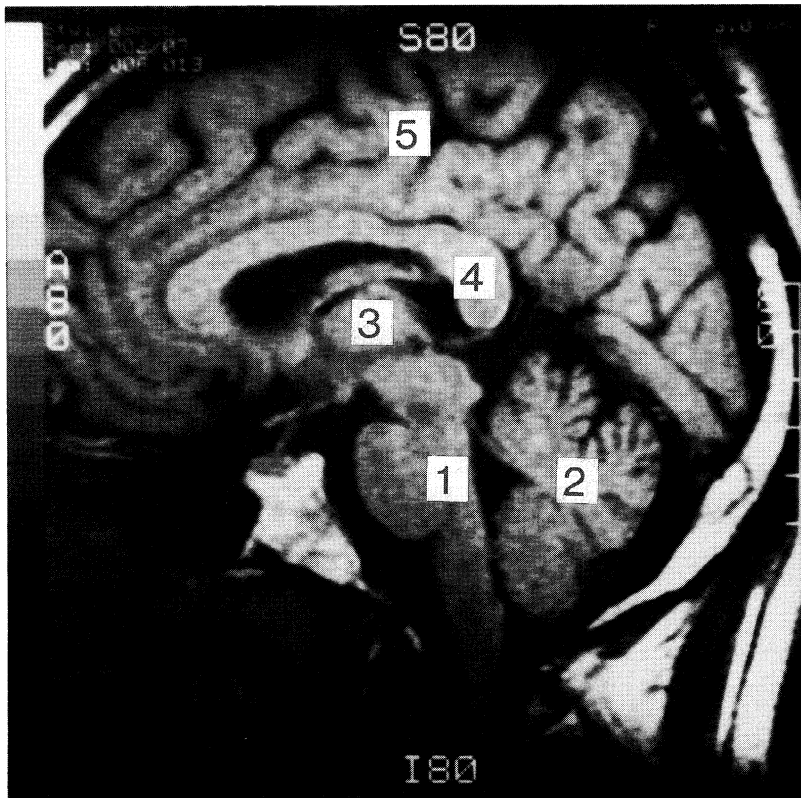


Figure 1

The brain as seen by magnetic resonance imaging. A lateral view of the inner surface of the brain is shown. Note the brain stem (1), cerebellum (2), thalamus (3), corpus callosum (4), and highly convoluted cerebral cortex (5).

muscles on the left side and receives visual information presented to the left of eye fixation. The partitioning of information to the right and left hemispheres is another example of the modularity or division-of-labor characteristic of brain function.

Perceptual Functions. The manner in which we perceive and act with people and objects in the world has typically not been the domain of intelligence. Perceptual functions are assumed to be rather constant and not indicative of our conceptual or intellectual abilities. However, these functions provide the basis for our conceptions and it is important to characterize how they are represented in the brain. Each sensory modality (e.g., touch, hearing, smell, taste, sight) is registered in different areas of the cerebral cortex. Touch information is first registered in the parietal lobe, a region on the lateral or outer sur-

face of the cerebral cortex. There is a topographic (i.e., body) representation of touch information in the brain; for example, nerve cells that register touch on the right thumb are located near nerve cells that register touch on the right index finger. Auditory information is first registered in the temporal lobe. It is initially represented tonotopically, that is, by sound frequency.

Sensory receptors for both smell and taste are chemical in nature. Molecules from the environment attach to chemoreceptors in the nose or on the tongue (see Carlson, 1991; Martin, 1989). In the cerebral cortex, smell information is first registered in the piriform cortex of the limbic lobe. The limbic lobe is a region of cortical cells located on the surface of each cerebral hemisphere near the center of the brain. Taste information is first registered in the insular cortex, a region in the frontal lobe. Both smell and taste information

subsequently converge onto another region in the frontal lobe, the orbitofrontal cortex, which perhaps explains why both of these senses contribute strongly to the perception of flavor.

Vision is the best studied sensory modality. It is first registered in the posterior area of the cortex known as the occipital lobe. Initially, visual information is represented spatially. When an individual suffers damage to the occipital lobe he or she will exhibit blindness in certain portions of visual space. If the damage is contained within a circumscribed area in one cerebral hemisphere, a scotoma or blind spot will be apparent in the visual field. If the damage involves much of the occipital lobe in one cerebral hemisphere, a hemianopia will occur in which the individual is functionally blind across an entire half field of vision. Interestingly, some visual information can be received by patients with occipital lobe lesions who are functionally blind. This phenomenon, known as blindsight, is exemplified by the ability to point to the locations of stimuli in the blind region, even though the individual claims to be blind and states that responses were based on mere guesses (Weiskrantz, 1986).

Language. Language was one of the first intellectual abilities to be localized in the brain. During the 1860s, Paul BROCA studied brain-injured patients with speech abnormalities (Schiller, 1992). He identified an area in the left frontal lobe that when damaged consistently led to a disturbance of speech production. The disturbance involved problems in articulation, word finding, and the formation of complete sentences. In these patients, speech was often slowed and telegraphic. However, language comprehension was not severely affected. Today, this disorder is often called Broca's aphasia, and the area in the left frontal lobe is called Broca's area.

Several years after Broca's discovery, Wernicke discovered another area in the left hemisphere that was important for language. Wernicke's area is in the temporal lobe and when damaged produces disturbances in language comprehension. Thus Wernicke's aphasia appears to be a problem in speech comprehension rather than speech production. Patients with Wernicke's aphasia speak fluently with good articulation, but they cannot understand speech and their own speech is not meaningful. Apparently, both the ability

to comprehend speech and the ability to arrange words into meaningful sentences are impaired in patients with Wernicke's aphasia. Wernicke developed a theory of language processing that suggested that speech sounds are first recognized and interpreted in the left temporal lobe and then, by axonal connections to Broca's area in the frontal lobe, are processed for speech production. Interestingly, Wernicke's area is adjacent to areas critical for the perception of basic, nonspeech auditory information. More recently, Geschwind (1972) developed a theory that elaborated Wernicke's theory and provided an anatomical basis for the connections between Wernicke's area and Broca's area.

These findings from aphasic patients suggest several important factors about the biological basis of verbal ability. First, the left hemisphere appears to play a prominent role in language ability. Specifically, the dominance of language in the left hemisphere occurs in approximately 96 percent of right-handed people and 70 percent of left-handed people. Other individuals tend to have a bilateral or right-hemisphere representation of language function. Second, the frontal lobe appears to be related to speech production, whereas the temporal lobe appears to be related to speech perception and comprehension. Third, language is not simply localized in a specific part of the left hemisphere, but various components appear to be distributed in different areas (e.g., Broca's area versus Wernicke's area). In normal language processing, these areas interact and work together to provide natural language functions.

Memory. In 1953, a patient known as H. M. underwent surgery for relief of severe epileptic seizures. The surgery involved bilateral excision of the medial temporal lobe, which included a large portion of the hippocampus, a structure that lies along the medial border of the temporal lobe. Following surgery, H. M.'s seizures were attenuated, but he exhibited a profound memory disorder—he was unable to remember events and information encountered since his operation. For example, a half hour after eating lunch, H. M. could not recall what he had eaten or if he had had lunch at all. H. M. was aware of his disorder and reflected upon his impairment as always “waking from a dream.” In other words, this disorder produced a

lack of continuity in the memory of events across time, even when the events were separated by only a few minutes. In fact, H. M. was not able to store facts and episodes that were encountered after his operation. Despite H. M.'s severe impairment in memory, there was no detectable impairment in perception, language, or other intellectual abilities. Moreover, H. M.'s memory for events prior to his operation were not severely affected. He was capable of recalling autobiographical episodes that occurred before his operation.

H. M.'s memory impairment—often called *organic amnesia*—affects information received from all sensory modalities and includes impairment of both verbal and nonverbal (e.g., spatial) memory (see Shimamura, 1992). For example, H. M. could not acquire new vocabulary words that had been added to the dictionary since his surgery (e.g., Jacuzzi). He also exhibited severe impairment on laboratory tests of word and picture recall, cued-word learning (e.g., learning word pairs), and multiple-choice, recognition memory. Despite the severity of his amnesia, H. M. was able to think and act normally. In fact, even some memory functions were spared, such as some functions of short-term memory, which is the ability to hold things in mind, such as a telephone number prior to dialing. However, as soon as information was out of conscious experience, it was forgotten. The analysis of H. M.'s amnesia stands as a milestone in our progress to understand memory in the brain. He provided the crucial evidence for the specific role of the medial temporal region in the process of memory formation and storage.

There are other neurological disorders that damage the medial temporal region and thus produce an amnesic syndrome similar to that seen in H. M. For example, tumors, head injuries, or vascular disorders (e.g., strokes) in this region can cause organic amnesia. Also, viral infections, ischemia (loss of blood flow to the brain), or hypoxia (loss of oxygen to the brain) can disproportionately affect the medial temporal region. In these disorders, organic amnesia is often the outstanding cognitive impairment. It appears that the hippocampus is particularly important for memory formation to occur. One theory for this is that the hippocampus connects or associates new incoming information to existing knowledge. It is known that this

brain region is intricately connected to a wide distribution of cortical areas (see Shimamura, 1992).

Spatial Ability. Although the left hemisphere predominates in the control of verbal ability, the right hemisphere appears to predominate in the control of spatial ability. Patients with right-parietal-lobe lesions have particular problems in organizing spatial relations (Kolb & Whishaw, 1990; Walsh, 1978). For example, these patients often have difficulty using maps and finding their way in unfamiliar places. They also have difficulty drawing objects, such as a house, in perspective. Patients with left-parietal-lobe lesions can also exhibit impairment in spatial ability, but spatial impairments following left-parietal lesions are typically not as severe as those following right-parietal lesions. Also, the pattern of impairment is often different. When asked to draw a house in perspective, a patient with a right-parietal lesion will fail to portray the figure in three-dimensions but may be able to draw some of the details (e.g., door, windows). A patient with left-parietal lesion typically will be able to construct the global features of the house and to draw the house with perspective, but will have problems drawing the details.

A particularly dramatic disorder of spatial ability can occur following right-hemisphere lesions. In this disorder, patients act as if they are not aware of the left side of the world, including the left side of their body. Such patients will fail to acknowledge individuals on the left side, fail to draw the left side of objects, and even fail to dress the left side of their bodies. This disorder is called *contralateral neglect syndrome* because damage to the right cerebral hemisphere causes the patient to neglect the opposite side of the world. This disorder is often associated with patients who recently have had a vascular stroke in their right hemisphere, typically in their parietal lobes. The symptoms are often transient and diminish following recovery from the stroke. *Contralateral neglect syndrome* can occur following damage to the left hemisphere, but much less frequently.

Contralateral neglect syndrome appears to be a problem in attending rather than perceiving. The patient can see objects in the neglected side, if asked to direct attention to the objects. One fascinating example concerned an Italian patient with *contralateral ne-*

glect syndrome who was asked to imagine himself standing at the Piazza del Duomo, a beautiful cathedral square in Milan (Bisiach & Luzzatti, 1978). When the patient described the view from the cathedral entrance looking out toward the square, he reported the details on the right side but neglected details on the left. However, when asked to imagine the view from the opposite end of the square and facing the cathedral, the patient now described all of the details that were previously neglected (i.e., information that was now on the right side). Thus, this patient was experiencing left-sided neglect of imagined scenes in addition to perceived scenes. The whole scene appeared to be accessible to the patient, but not all at once. Only half of the scene was available or could be attended from any imagined vantage point.

Executive Functions. How do you know that something is on the “tip of your tongue” when you cannot retrieve the information? How do you know when to stop searching for an item in memory? How do you know that the average man’s tie is less than 10 feet long? These questions address the ability to monitor and control your thought processes, or what are called executive functions. Planning, problem solving, reconstructing memories, and making rational decisions all require the use of executive functions. This is the essence of what we call intellectual reasoning. The frontal lobes of the brain appear to mediate executive functions. The frontal lobes encompass roughly a third of the cerebral cortex in humans. They provide the main cortical output for coordinating and executing muscle movements (Luria, 1973). Also, Broca’s area, in the frontal lobes, contributes to the coordination and execution of speech and language. Other areas in the frontal lobes appear to contribute to the coordination and execution of thought processes.

The frontal lobes are intricately connected to many other areas in the cerebral cortex. As a result of these extensive connections, damage to this area is associated with a variety of mental dysfunctions in humans, including disorders of motor control, personality, emotion, language, problem solving, and memory. Studies of brain-injured patients suggest that different regions within the frontal lobes mediate different executive functions. In terms of intellectual function, patients with frontal-lobe lesions are typically impaired when asked to plan strategies or to determine methods

of approaching a problem. They appear to be “stimulus bound” and act without forethought, and thus fail to respond appropriately to questions that require deliberate thought or the formation of inferences. For example, few would know specifically the length of the average man’s tie, but most would be able to estimate or infer the approximate length. Similarly, if strategies to access your library of memories were not well planned, it would be difficult to know whether you were close to the retrieval of some fact or whether you were far from retrieval. These kinds of errors are common in patients with frontal lobe damage.

There have been a variety of theories to explain the array of mental disorders associated with damage to the frontal lobes. Some researchers have suggested that frontal-lobe function is what makes humans unique. In particular, it has been considered to be the basis for high-level, intellectual reasoning and abstraction (Goldstein, 1939; Halstead, 1947). Although frontal-lobe function does play a role in intellectual reasoning, such as decision making and problem solving, intelligence appears to involve an enormous number of mental components distributed throughout the brain (e.g., perceptual analysis, language, spatial ability), and any number of these components may be particularly important for a specific intellectual task. In fact, Donald Hebb (1945) argued strongly against the view that the frontal lobes affect intelligence, and he demonstrated that frontal-lobe damage in humans did not grossly affect IQ scores. Recent findings suggest that patients with frontal-lobe damage do not demonstrate significantly impaired IQ scores because IQ measures primarily test knowledge and the ability to solve simple problems, but they do not adequately test executive function.

One possibility is that the frontal lobes serve executive function by organizing and integrating mental components or processes. In other words, the frontal lobes may work as a central executive or supervisor of the various specific components involved in language, memory, and spatial ability (see Baddeley, 1986). This view would explain the importance of the frontal lobes on tasks that involve planning and monitoring thoughts and actions. It also explains why patients with frontal-lobe damage appear to act impulsively or without forethought. That is, they often behave according to what pops into mind without filtering or

censoring inappropriate actions or behaviors. In general, patients with frontal-lobe damage often appear to have no inhibitory control.

CONCLUSION

Based on this brief summary of the ways in which intellectual function is mediated by brain mechanisms, it is apparent that several basic findings have been established. First, different parts of the cerebral cortex contribute to different aspects of intellectual function. Second, brain damage can lead to a disruption of some aspects of intellectual function while leaving other aspects entirely intact. Third, general intelligence cannot be attributed to any single area of the brain; it clearly occurs as a result of a concerted effort of many specialized processors. Finally, the frontal lobes appear to act as a supervisor or executive of mental activity and appear to be involved in the ability to plan, monitor, and organize thoughts and actions.

One theory of intelligence incorporates these findings from neurological patients. Howard Gardner (1983) developed a theory of "multiple intelligences" in which he suggests that there is not any single or general criterion for intelligence. Instead, Gardner suggests that intelligence is based on multiple components and individuals may excel in one or many of these components. Each component refers to a different mental ability, such as linguistic, musical, logical-mathematical, spatial, bodily-kinesthetic. Excellence in one form of intellectual function does not necessarily imply excellence in others. However, excellence in many of these functions may reflect a high level of intellectual capacity. Thus, just as findings from brain-injured patients indicate that there are multiple forms of mental deficiency, Gardner's theory of intelligence suggests that there are multiple forms of mental proficiency (see MULTIPLE INTELLIGENCES THEORY).

The theory of multiple forms of intelligence may not encompass the entire realm of intelligence capacity. As noted by Galton (1883) and by others, there may be aspects of general neuronal efficiency that may make every component of intelligence work in a more effective manner. Moreover, a more effective supervisor (i.e., frontal lobe) could facilitate the organization and integration of intellectual functions. Therefore, the view of multiple intellectual components must be tem-

pered by the notion of general neuronal efficiency that could affect all components. This more global view would account for the finding that proficiency of intellectual functions (e.g., linguistic, musical, logical-mathematical) is often correlated in highly intelligent individuals. These struggles about the notion of intellectual capacity can be approached effectively by investigations of brain function. Further analyses of these brain functions will undoubtedly lead to a more complete understanding of intelligence. [This work was supported by a grant from the National Institute on Aging (AG09055).]

(See also: BRAIN, PATHOLOGIES OF; EEG EVOKED POTENTIAL.)

BIBLIOGRAPHY

- BADDELEY, A. (1986). *Working memory*. Oxford, UK: Oxford University Press.
- BIRREN, J. E., & WELFORD, A. T. (EDS.). (1965). Age changes in speed of behavior: Its central nature and physiological correlates. *Behavior, aging, and the nervous system*, (pp. 191–216). Springfield, IL: Charles C. Thomas.
- BISIACH, E., & LUZZATTI, C. (1978). Unilateral neglect of representational space. *Cortex*, 14, 129–133.
- CALVIN, W. H. (1990). *The ascent of mind*. New York: Bantam Books.
- CARLSON, N. R. (1991). *Physiology of behavior*. Boston: Allyn & Bacon.
- GALTON, F. (1883). *Inquiries into human faculty and its development*. London: Macmillan.
- GARDNER, H. (1983). *Frames of mind*. New York: Basic Books.
- GESCHWIND, N. (1972). Language and the brain. *Scientific American*, 226, 76–83.
- GOLDSTEIN, K. (1939). Clinical and theoretical aspects of lesions of the frontal lobes. *Archives of Neurology and Psychiatry*, 41, 865–867.
- GOULD, S. J. (1981). *The mismeasure of man*. New York: W. W. Norton.
- HALSTEAD, W. C. (1947). *Brain and intelligence*. Chicago: University of Chicago Press.
- HEBB, D. O. (1945). Man's frontal lobes: A critical review. *Archives of Neurology and Psychiatry*, 54, 10–24.
- KOLB, B., & WHISHAW, I. Q. (1990). *Fundamentals of human neuropsychology*. New York: W. H. Freeman.

- KOSSLYN, S. M., & KOENIG, O. (1992). *Wet mind: The new cognitive neuroscience*. New York: Free Press.
- LURIA, A. R. (1973). *The working brain*. New York: Basic Books.
- MARTIN J. H. (1989). *Neuroanatomy*. Norwalk, CT: Appleton.
- MATARAZZO, J. D. (1992). Psychological testing and assessment in the 21st century. *American Psychologist*, 47, 1007–1018.
- SCHILLER, F. (1992). *Paul Broca*. New York: Oxford University Press.
- SHIMAMURA, A. P. (1992). Organic amnesia. In L. Squire (Ed.), *Encyclopedia of Learning and Memory*. New York: Macmillan.
- SPEARMAN, C. (1923). *The nature of "intelligence" and the principles of cognition*. London: Macmillan.
- WALSH, K. W. (1978). *Neuropsychology: A clinical approach*. Edinburgh: Churchill Livingstone.
- WEISKRANTZ, L. (1986). *Blindsight: A case study and implications*. Oxford: Oxford University Press.

ARTHUR P. SHIMAMURA
JOAN W. SHIMAMURA

BRAIN, PATHOLOGIES OF THE Structural alterations of the human brain may be produced by a variety of factors—among them trauma, vascular disease, toxins, infectious agents, and normal aging processes. Depending on their extent, site, and momentum, as well as other variables, such as compensatory reactions and the age of the individual, these structural alterations may interfere with the brain mechanisms that mediate cognitive functions. When such interference does occur, intellectual impairment in one form or another is an inevitable consequence.

Classification of these structural changes proceeds from various standpoints, such as a disease picture (e.g., Alzheimer's disease), an obvious deviation (e.g., congenital malformations), a cause (e.g., trauma or infection) or solely an interest in changes in the brain over the life span (e.g., aging). Differing factors may produce essentially the same structural alterations, and there is overlap between age-associated normal and pathological changes. Recently, advances in the understanding of the mechanisms underlying transmission of information in the nervous system have created a func-

tional neuropathology, in which the focus of interest is on the disturbances in neurotransmitter operations that lead to significant cerebral dysfunction and resulting cognitive impairment.

METHODS OF STUDY

The time-honored technique for bringing to light gross pathological changes in the brain, such as tumors, hemorrhages, and atrophy, is the autopsy. Attempts to correlate autopsy findings with patients' deranged functions while they lived (including specific cognitive impairments) achieved notable success in the latter decades of the nineteenth century. During this period, among other associations, nonfluent and fluent aphasic disorders (quality of pronunciation) were related respectively to frontal and posterior temporal lesions of the brain's left hemisphere; visual object agnosia (loss or impaired recognition of familiar objects) was related to bilateral occipital lobe disease; and a peculiar assemblage of cognitive, emotional, and personality changes was found to be so closely associated with frontal lobe disease that it was given the designation "frontal lobe syndrome." The rise of microscopic anatomy made possible histological study of the brain, which disclosed the distinctive cellular changes characteristic of (and indeed defined) specific degenerative diseases, such as Alzheimer's disease, multi-infarct dementia, and the various types of tumors. Histological quantitative study paved the way for estimating the degree of cell loss in old age and dementing diseases; for determining whether grossly perceived "atrophy" did or did not reflect actual cell loss; and for the minute examination necessary for elucidating the status of neurotransmitter functions.

During the early and middle decades of the twentieth century, a variety of techniques, including skull x-ray, pneumoencephalography, ventriculography, cerebral arteriography, and electroencephalography, were developed to visualize structural or functional abnormalities in the brain. These procedures were of tremendous value for neurological and neurosurgical diagnosis; however, because they provided only imprecise indications of the locus (place) and extent of lesions, they had limited significance for the study of specific brain-behavior relationships in living patients. Reports of surgical excisions and plotting the paths of

penetrating brain wounds were more informative but could also be quite misleading.

Beginning in the 1970s, the successive introduction of new techniques of elucidating cerebral structure and function in the living patient led to a revolution in the study of brain-behavior relationships. Some techniques, such as X-ray computed tomography (CT) and nuclear magnetic resonance imaging (MRI), visualized structural changes in the brain with remarkable precision. For example, the utilization of CT or MRI enables the neuroradiologist to specify with a considerable degree of confidence that a lesion is situated in the territory of the left angular gyrus or on the inferior surface of the right occipital lobe. Three-dimensional mapping of lesions from MRI data in the living patient is a recent development of extraordinary importance. This technique generates a picture of the location and extent of a lesion that is comparable to, and indeed even more informative than, autopsy findings (Damasio & Frank, 1992).

Other techniques, such as the measurement of regional cerebral blood flow (rCBF), positron emission tomography (PET), and the recording of evoked potentials (EP) that are indicative of the level of neural activity, make it possible to observe the changes in brain function that occur during the course of a subject's performance. For example, cerebral blood-flow studies of normal subjects have found that verbal-task performance generates greater activity in the left hemisphere than in the right and, conversely, that visuospatial-task performance generates greater activity in the right hemisphere. People who have sustained a unilateral stroke show reduced activation in the affected hemisphere during the course of task performance—and a consequent loss of activity on the opposite side of the body (since one side of the brain controls the opposite side of the body). Utilization of PET scans to plot the lesions producing aphasia has demonstrated that the area of reduced functional activity is significantly larger than the structural alteration indicated by CT scans.

TYPES OF PATHOLOGY: COGNITIVE CORRELATES

Degenerative Diseases. Degenerative diseases result from progressive pathological changes in various

regions of the brain or spinal cord that primarily involve neural tissue. Genetically or environmentally determined errors in metabolism are the primary causative factors underlying these disorders. Although considerable variation exists in the profile of cognitive disabilities associated with degenerative diseases of the brain, the diseases generally produce the pervasive decline in level of intellectual functions denoted by the nonspecific term *dementia*.

The most common form of dementia is Alzheimer's disease, which, for the most part, affects individuals who are 60 years of age or older and hence is said to be "age-associated." From a neuropathological standpoint, the disease is defined by certain specific changes (neurofibrillary tangles, senile plaques) in the brain substance that are disclosed by postmortem histological examination. Since a definitive diagnosis rests on autopsy findings, the more cautious terms, *probable Alzheimer's disease* and *dementia of the Alzheimer type*, are often used in diagnosing the condition in the living patient.

Alzheimer's disease is characterized by an extensive loss of neurons, resulting in widespread atrophy of the brain. The atrophy, which is clearly seen in enlargement of the cerebral ventricles and in widening of the sulci and thinning of the gyri of the cerebral cortex, is bilateral and often more or less symmetrical. Within this setting of widespread atrophy, a predilection for particularly severe involvement of temporal and frontal lobe cortex exists, with sparing of the parietofrontal sensorimotor areas and the occipital cortex. Pronounced involvement of the parieto-occipital cortex or of Wernicke's (speech) area in the temporal lobe is seen in some cases, however.

The behavioral picture of Alzheimer's disease is one of an insidious onset of cognitive decline and personality change. The *modal* pattern of cognitive impairment is characterized by prominent defects in short-term and recent memory, reasoning ability, and practical judgment, with sparing of motor skills, sensory capacities, and basic speech functions. Personality changes include emotional lability (instability), intractable stubbornness, and childish egocentricity. "Atypical" presentations, however, in which the onset of this illness is signaled by frank aphasic disorder or pronounced impairment of visuo-perceptual functions, are not rare.

Numerous studies have found correlations of moderate size between the severity of cognitive impairment in Alzheimer patients and the degree of neuropathological involvement, as reflected in the brain—ventricular enlargement, depth and widening of cortical sulci, and counts of neurofibrillary tangles and senile plaques. Similar associations have been found between cognitive impairment and functional abnormality as assessed by measurement of cerebral blood flow and PET scan. Moreover, specific associations between discrete cognitive deficits and regional neuropathology have also been described (e.g., between aphasic disorder and involvement of Wernicke's area and between visuospatial disability and right-hemisphere pathology).

Progress in knowledge of the neuropathological correlates of Alzheimer's disease has been particularly rapid in the 1980s. As new techniques of investigation are more widely applied, important advances in understanding may be expected. A noteworthy recent finding is that loss of cortical dendritic connections (and not neurofibrillary tangles or senile plaques) is the variable that is most closely associated with the severity of cognitive impairment in Alzheimer patients (Terry et al., 1991). If this finding is confirmed, it may lead to a basic redefinition of the essential neuropathology of the disease.

A number of other less frequently occurring forms of degenerative disease, such as Pick's disease (a circumscribed atrophy of the brain usually affecting the temporal or frontal lobes), present a distinctive neuropathological picture but cannot always be differentiated from Alzheimer's disease on clinical and behavioral grounds. Cognitive changes have been described in association with Parkinson's disease, which involves degeneration of the basal ganglia and which is characterized clinically by tremor, slowness in movement, and motor rigidity. Frontal lobe degeneration is a recently described form of dementia (Neary & Snowden, 1991). The disorder is characterized pathologically by an atrophy of the frontal and anterior temporal lobes that is different from that found in other degenerative diseases. The behavioral picture is that of a pronounced frontal lobe syndrome, such as might be seen in the advanced stage of a frontal lobe tumor or a severe traumatic brain injury, with striking

changes in personality as well as cognitive defects. Lack of concern, neglect of personal responsibilities, and disinhibition are prominent features of behavior. Neuropsychological assessment brings to light reduced speech production without frank aphasia, perseveration (repetition of words), and motor impersistence with sparing of visuo-perceptual and visuoconstructive (perception of forms and constructional) abilities and recent memory.

Aging. To regard old age as a disease might well be considered an affront to human dignity but, from the standpoint of neuropathology, it *is* a disease. The same changes in the brain that are seen in typical Alzheimer's disease—atrophy, ventricular enlargement, widening of the cortical sulci, neuritic plaques, and dendritic loss—are found in the brains of normal elderly persons, albeit to a much lesser degree. Moreover, the intensity of these changes increases progressively with advancing age—the typical findings being that the correlation coefficients between such variables as ventricular size, cortical gyral atrophy, and incidence of neuritic plaques and age are small but statistically significant.

The decline of diverse cognitive performances with age, particularly on tasks making demands on rate of information processing and on the capacity for learning and retention, is well documented. It is reasonable to expect that there is a relationship between the state of the brain and cognitive efficiency in normal aging. The findings of the few studies that have been done to date are equivocal because of methodological limitations. There are sensory changes with age, such as diminished visual acuity, contraction of the visual field, and hearing loss, which may affect performance on cognitive tasks; these changes must be taken into account in relating performance to the status of the brain. Moreover, educational and occupational background is closely correlated with level of performance and this also needs to be controlled before the "pure" association between cognitive ability and brain status is assessed.

A point of some interest is that the resting rate of cerebral blood flow (CBF) declines with advancing age, thus shifting in the direction of the reduced CBF found in patients with cerebrovascular disease and Alzheimer's disease. The decline is most rapid in the pre-

frontal region and less rapid in those regions that are activated by continuous daily activity, that is, the left precentral and perisylvian areas (speech) and the right parietal area (visuospatial activity). At the same time, it has been shown that, although older normal subjects have lower resting CBF values than younger subjects, performance on cognitive tasks activates CBF to the same degree in both groups. Taken together, these observations suggest that sustained engagement in high-level cognitive activity may maintain CBF levels in less frequently activated regions and thus retard age-related structural changes in the brains of elderly persons. "Conversely, inactivity associated with social withdrawal, depression, or premature retirement may hasten cerebral deterioration in the aged, and this has long been suspected, based on clinical observation" (Meyer & Shaw, 1984).

Cerebrovascular Diseases. Cerebrovascular diseases occur in a variety of forms and range widely in severity, from occlusion of a small blood vessel that produces temporary cerebral dysfunction (transient ischemic attack) to widespread involvement of the cerebrovascular system resulting in marked impairment of cognitive function (multi-infarct dementia).

The most common form of cerebrovascular disease is stroke, a sudden onset of brain dysfunction produced either by occlusion (stoppage of the flow of blood) in a blood vessel or by hemorrhage. Occlusion, the more frequent of the two mechanisms, deprives the area of the brain supplied by the blood vessel of the oxygenated blood required for its functioning and survival (ischemia) and eventually results in destruction (infarction) of that area.

Cerebrovascular disease, specifically the focal ischemic infarct, has been a most valuable "experiment of nature" for the investigation of brain-behavior relationships; hence, it is of particular interest to the neuropsychologist (Benton, 1991). Practically all our knowledge of aphasic (speech) disorders is based on the study of patients with well-defined, precisely localized ischemic infarcts. The neurological basis of specific cognitive and perceptual deficits, such as loss of the ability to recognize faces (prosopagnosia) or familiar objects (visual object agnosia) and acquired color blindness (achromatopsia), has been elucidated by autopsy or neurodiagnostic examination of poststroke

patients. Moreover, it has been possible to relate distinctive combinations of such cognitive deficits to circumscribed ischemic lesions of the brain (Benton, 1991, 1992).

Dementia, resulting from widespread or particularly severe cerebrovascular disease, tends to differ in character from that seen in Alzheimer's disease. Frank aphasic disorder and dysarthria (impairment in pronunciation) are encountered far more frequently in patients with multi-infarct dementia than in Alzheimer patients. Other specific deficits indicative of focal brain disease, such as poor perception of forms and poor constructional ability, are also more common in vascular dementia. Considerable variability exists in the behavior pictures in the category of multi-infarct dementia. The essential reason for this variability is that the underlying neuropathology differs from one person to another, with primary involvement of diverse sites in the brain including the cerebral cortex, the underlying white matter, the basal ganglia, or the thalamus.

Congenital Malformations. Congenital malformations reflect abnormal prenatal development caused either by genetic factors or by environmental agents, such as toxins and infections. Examples are lack of development (agenesis) of the corpus callosum, fetal alcohol syndrome (a defective infant of an alcoholic mother), and congenital rubella (German measles) infection, all of which involve varying degrees of cognitive impairment. Numerous types of congenital malformation vary widely in extent and severity, for example, from anencephaly (absence of the cerebral hemispheres) to pencephaly (a limited focal defect in the structure of one brain hemisphere). Hydrocephalus—an excessive accumulation of cerebrospinal fluid in the cerebral ventricles that can destroy brain tissue—may be produced by a variety of congenital malformations.

Many congenital malformations are accompanied by severe and pervasive mental deficiency. Other conditions, in which the cognitive outcome is variable, have been of greater interest to psychologists. One such condition is agenesis of the corpus callosum (lack of development of the massive band of fibers that provide for transmission of information between the two hemispheres). Surgical cutting of the corpus callosum in adult patients (undertaken for the relief of intrac-

table epilepsy) leads to a series of well-defined cognitive changes that can be related directly to the loss of the connection between the hemispheres. The behavioral correlates of callosal agenesis are quite different. The resulting overall intellectual level varies considerably, ranging from frank mental retardation to high average capacity with the median level being dull average. Acallosal subjects show very few of the performance characteristics indicative of defective interhemispheric integration that are seen in patients with surgical transections. The probable reason for this difference is that during the course of development, they have learned to use other interhemispheric pathways to achieve integration.

Another condition of interest is congenital or early acquired hydrocephalus, which, as has been mentioned, is not itself a congenital malformation but rather the product of one. The excessive accumulation of cerebrospinal fluid in the ventricles with resulting increased intracranial pressure leads to atrophy (death) of surrounding brain tissue, sometimes to such an extreme degree that the cerebral cortex is found to be paper thin.

The determinants of the resulting intellectual outcome in congenital or early acquired hydrocephalus are not well understood. A variety of factors—brain mass, thickness of the cortical mantle, the nature of the congenital structural abnormality responsible for the hydrocephalus, the presence of other congenital anomalies, general health status and the occurrence of seizures, whether or not a shunt operation has been performed to relieve the excessive intracranial pressure, socioeconomic status, and parental concern all appear to play a role in affecting the mental development of the hydrocephalic child. Thickness of the cerebral mantle of less than 1 cm is almost always associated with frankly subnormal intelligence, but even in this case some striking exceptions have been reported. A hydrocephalic child may appear to develop normally during the first 2 or 3 years of life and then manifest cognitive disabilities as he encounters new task demands. The type of malformation responsible for the hydrocephalus appears to be of importance. Total brain mass is a correlate of intellectual level but the size of the relationship is moderate. Nonverbal test performances (such as copying a design) is

likely to be poorer than verbal skills (such as vocabulary level). Eye-hand coordination, in all probability associated with basic visual defects, is likely to be poor. The extreme variability in outcomes that range from those who are severely retarded to those who are intellectually gifted, coupled with the limited implications of empirical studies to date, have left most of these issues unsettled, however. Comprehensive multivariate studies should go far toward resolving many questions (Donders, Canady, & Rourke, 1990).

Traumatic Brain Injury. There are two types of traumatic brain injury: (1) penetrating head injuries (such as are produced by gunshot wounds); and (2) closed head injuries (such as in an automobile accident), in which the brain is injured by forces acting upon it through the intact skull. The effects of these two types of injury are rather different.

Penetrating Head Injury. Direct gunshot wounds to the head are associated with a high frequency of mortality. If the victim survives, however, the passage of a high-speed missile through the brain is likely to have produced a focal lesion of limited size. This is especially true of wartime head injuries, where the missile is more often than not a fragment of an exploded bomb or shell rather than a bullet. These patients have provided a valuable opportunity for the study of specific human brain-behavior relationships because they are likely to show specific cognitive deficits within the context of relatively spared "general intelligence" (Newcombe, 1969). For example, studies of these patients were the first to furnish strong evidence of the dominance of the right hemisphere for visual and spatial abilities. Later analyses disclosed that wounds in the upper parietal region produced visuospatial (recognition of location) defects, while those in the inferior occipitotemporal area led to impairment in the recognition of faces. Similarly, it was found possible to relate isolated aphasic defects such as loss of the ability to name objects or to write words to small circumscribed lesions in the left hemisphere.

Closed Head Injury. Although closed head injury often produces small areas of brain damage in the form of bruises or blood clots, its primary effect is diffuse injury of varying severity involving shearing of nerve fibers and brain swelling (Levin, Benton, & Grossman, 1982; Levin, Eisenberg, & Benton, 1989). In addition,

there may be late aftereffects such as enlargement of the ventricles (indicating brain atrophy) and progressive degeneration of nerve fibers. Loss of consciousness following the trauma is a characteristic occurrence; in the majority of cases it is of short duration but prolonged loss of consciousness (coma) is not uncommon. Following coma, there is often a period of posttraumatic amnesia, during which the patient, although conscious and reactive, does not remember successive ongoing events from one occasion to another. Since both the duration of coma and of posttraumatic amnesia have been found to be significantly predictive of the final behavioral outcome of a closed head injury, they are frequently utilized as indices of the severity of injury.

As would be expected, the cognitive outcome of closed head injury is quite variable, depending on a number of factors including the severity of injury, the region of the brain most affected, the age of the patient, the availability of appropriate medical and surgical care, and the presence of a social support network. The frontal region of the brain is very often the site of contusions, hematomas, and intracranial hemorrhages. As a consequence, a clinical picture of moderate frontal lobe disease with its characteristic cognitive and personality changes is commonly seen. The major changes are disturbances of attention and concentration, slowed information processing, and impaired learning and retention. Frank aphasic disorder and visuospatial disability, such as are shown by patients with penetrating brain wounds, are rare. Personality changes in the form of emotional instability and impulsive behavior are quite prominent. The cognitive deficits of these frontally injured patients are not adequately reflected in performance on conventional intelligence test batteries that so often include measures of "crystallized" intelligence, such as vocabulary level, range of information and social judgment. Hence, total scores on such an intelligence test may show little decline. The disparity between an intelligence score within the normal range and the patient's obvious behavioral incompetence is sometimes a source of puzzlement to family members and friends. Special test methods can bring the impairment in attention, learning, and reasoning capacity into sharp focus (Levin, Eisenberg & Benton, 1991).

Postconcussion syndrome or *posttraumatic symptom-complex* are terms that refer to the physical and psychological complaints (e.g., headache, fatigability, poor concentration and memory, anxiety, insomnia) voiced by patients who have sustained only a mild head injury with a brief period of unconsciousness and who otherwise appear to have no significant neurological or cognitive deficits. There has been a longstanding controversy between those clinicians who see the postconcussion syndrome as psychological in origin (if not actual malingering) and those who contend that the patient's complaints have an organic basis. The thrust of recent investigations has been to support the belief that more often than not the syndrome is in fact related to subtle brain pathology (Levin, Eisenberg & Benton, 1989).

Tumors. The majority of intracranial tumors involve the neuroglia (the network of cells supporting neural tissue) and the meninges (the membrane covering of the brain) and not the brain tissue itself. These new tumor growths (neoplasms) influence cerebral structure and function by exerting excessive pressure (often through the accumulation of fluid) on neural tissue and this may lead to displacement and destruction of brain structures. Given this mass effect, the behavioral consequences of neoplastic disease on the individual are likely to be of a general rather than focal nature with lack of spontaneity and initiative, inability to concentrate, indifference, and forgetfulness as prominent features. Tumors in the prefrontal region lead to the cognitive and personality changes of the frontal-lobe syndrome described above. Tumors in other regions of the brain can also produce this syndrome if the mass swelling and pressure effects are sufficiently strong.

The behavioral consequences of tumors depend very much on their age and rate of progression. An early tumor may be behaviorally "silent" (i.e., have no effects), and in some locations even a large tumor may not exert significant mass effects. Due to early diagnosis, the devastating behavioral abnormalities associated with tumors are less often reported today than in earlier times. A recent comparative study of matched groups of tumor and stroke patients found that the stroke patients were cognitively impaired far more seriously than were the tumor patients. Indeed, the per-

formances of some tumor patients were quite normal on a lengthy battery of verbal and nonverbal tests (Anderson, Damasio, & Tranel, 1990).

The Neuropathology of a Learning Disability. Developmental dyslexia (failure to learn to read) and the diverse hypotheses—cognitive, physiological, sociocultural—that have been advanced about its nature and causation are familiar to psychologists and educational specialists. Autopsy examination of the brains of dyslexic adults has generated some interesting and unexpected findings and provided yet another hypothesis to account for the disorder. Galaburda (1989) and his coworkers have identified cellular abnormalities throughout the cortex, and particularly in the language-associated perisylvian area of the left hemisphere, in the brains of a number of cases. Moreover, the anatomical asymmetry of the two hemispheres that is found in the brains of normal individuals was not apparent in the brains of the dyslexic cases.

EPILOGUE

This entry has dealt with the major types of human brain pathology that disrupt cognitive functions. The reader will have noted that the relationships that have been established are of a simple associative nature, for example, between left-hemisphere damage and aphasic disorder, between right-hemisphere damage and visuospatial disability, between frontal-lobe damage and deterioration of personal conduct. The observed damage presumably deranges the brain mechanisms that mediate these behavioral functions. The identification of these mechanisms is the task of the newer functional neuropathology.

(See also: BRAIN; BROCA, PAUL.)

BIBLIOGRAPHY

- ANDERSON, S. W., DAMASIO, H., & TRANEL, D. (1990). Neuropsychological impairments associated with lesions caused by tumor or stroke. *Archives of Neurology*, *47*, 397–405.
- BENTON, A. L. (1991). Cerebrovascular disease in the history of clinical neuropsychology. In R. A. Bornstein & G.

- Brown (Eds.), *Neurobehavioral aspects of cerebrovascular disease*. New York: Oxford University Press.
- BENTON, A. L. (1992). Gerstmann's syndrome. *Archives of Neurology*, *49*, 445–447.
- DAMASIO, H., & FRANK, R. (1992). Three-dimensional in vivo mapping of brain lesions in humans. *Archives of Neurology*, *49*, 137–143.
- DONDERS, J., CANADY, A. I., & ROURKE, B. P. (1990). Psychometric intelligence after infantile hydrocephalus. *Child's Nervous System*, *6*, 148–154.
- GALABURDA, A. M. (1989). Ordinary and extraordinary brain development. *Annals of Dyslexia*, *39*, 67–89.
- LEVIN, H. S., BENTON, A. L., & GROSSMAN, R. G. (1982). *Neurobehavioral consequences of closed head injury*. New York: Oxford University Press.
- LEVIN, H. S., EISENBERG, H. M., & BENTON, A. L. (1989). *Mild head injury*. New York: Oxford University Press.
- LEVIN, H. S., EISENBERG, H. M., & BENTON, A. L. (1991). *Frontal lobe function and dysfunction*. New York: Oxford University Press.
- MEYER, J. S., & SHAW, T. G. (1984). Cerebral blood flow. In M. L. Albert (Ed.), *Clinical neurology of aging*. New York: Oxford University Press.
- NEARY, D., & SNOWDEN, J. S. (1991). Dementia of the frontal lobe type. In H. S. Levin, H. M. Eisenberg, & A. L. Benton (Eds.), *Frontal lobe function and dysfunction*. New York: Oxford University Press.
- NEWCOMBE, F. (1969). *Missile wounds of the brain*. London: Oxford University Press.
- TERRY, R. D., MASLIAH, E., SALMON, D. P., BUTTERS, N. et al. (1991). Physical basis of cognitive alterations in Alzheimer's disease: Synapse loss is the major correlate of cognitive impairment. *Annals of Neurology*, *30*, 572–580.

ARTHUR BENTON

BRIGHAM, CARL C. (1890–1943) Carl Campbell Brigham was born in 1890 in Marlboro, Massachusetts, the son of privileged descendants of Mayflower families, and he died in 1943 at age 53. After a short and apparently unsatisfactory period at Harvard, he transferred to Princeton, where he earned his bachelor's (1912), master's (1913), and doctoral (1916) degrees in psychology. Very early in his career he showed a clear interest in the emerging field of

intelligence testing; in 1917, a year after his receipt of the doctorate, and eager to contribute to the war effort even before the United States's actual involvement in the war, he went to Canada to serve as a psychologist in the veterans' reeducation division of the Canadian Military Hospitals Commission. Apparently, Brigham had already come to be known as a talented young man, and shortly after his arrival in Canada, he was visited by Robert M. Yerkes, then president of the American Psychological Association. Yerkes wanted to discuss with Brigham the role he might take, with other U.S. psychologists, in the war effort—especially in assessing the intelligence of U.S. recruits. This meeting marked the beginning of a long, close association between the two men, in which Brigham, the protégé, seemed to benefit from the philosophies and research interests of Yerkes, the mentor.

When the United States entered World War I, Yerkes was appointed head of Army psychology and in that capacity organized the first application of group administered multiple-choice standardized intelligence tests on a large scale. These Army tests, which were based principally on the earlier 1916 Stanford-Binet tests, were designed to help select recruits for officer training and had been adapted to enable group administration for that purpose by Yerkes, Lewis M. TERMAN, and Arthur S. Otis. Two types of tests were developed: the Army Alpha tests, designed for literate Army recruit examinees, and the Army Beta tests, which depended only on oral and gestured instructions, designed for illiterate examinees and for those who could not read English. Urged by Yerkes to join him in his Army work, Brigham enlisted in July 1917, and shortly thereafter he was administering preliminary tests at Camp Dix, New Jersey, where he collected the data that formed the basis of the research he later reported in *A Study of American Intelligence* (1923). (See also ARMY ALPHA AND BETA TESTS OF INTELLIGENCE.)

Collaborating with Yerkes and other Army psychologists (among them, E. L. THORNDIKE, L. M. Terman, Arthur S. Otis, Frederick L. Wells, and Walter Dill Scott), Brigham exerted an influential role, not only in this particular testing program, but in the development of the short-answer, and later, the multiple-choice tests of the future.

In 1920, Brigham returned to Princeton as an assistant professor (in 1924 he became an associate professor and in 1928 a full professor) and continued to work on the Army Alpha tests. Brigham's interest in these tests led him to become the secretary of Princeton's committee on admissions, on which he served for many years.

In 1923, the College Board, in an effort to identify promising students for whom the existing free-answer achievement tests were not wholly adequate for this purpose, commissioned Brigham to chair a committee charged with the design and development of a "psychological examination." This was to be appropriate for all students preparing to enter college, irrespective of field of study, and it was to help the colleges assess the student's intelligence and aptitude for academic work. The first Scholastic Aptitude Test (SAT) was developed in 1925 and administered in 1926. Scores on the SAT were intended for use by college admissions officers as a supplement to the scores provided by the subject matter achievement tests (see SCHOLASTIC ASSESSMENT TEST).

It was in his scientific orientation to testing and test development in particular and his empirical approach to the refinement of the testing instruments that Brigham made his more important and lasting contributions. Perhaps more fundamental was his leadership in making a break with a deeply held philosophy and with assumptions concerning the intellect shared by many of the influential psychologists at the time, some of them active and dedicated members of the U.S. eugenics movement.

It should be recalled that in the United States, the early decades of the twentieth century were marked by a generally antipathetic, even repressive, attitude shared by many members of the Protestant "Nordic" establishment toward all other ethnic, racial, and religious groups—Asians, eastern and southern Europeans, Catholics, Jews, and especially blacks; the period from the late nineteenth century through the early twentieth century was notorious for vicious racism in the United States (see Franklin, 1965, pp. 431–437). Undoubtedly, this social climate was part of the attitudinal legacy that Brigham inherited. On the scientific side, it should also be recalled that the notion of intelligence as a unitary trait, passed on genetically from

parent to child and relatively impervious to environmental effects, enjoyed considerable currency—much more during the early decades of the twentieth century than now. This view, held by many of the testers at the time, including Yerkes, Terman, Brigham, and others, coupled with the naive view—perhaps characteristic of an overconfident young science—that the intelligence test was an unerring indicator of native ability, led Brigham to draw some conclusions that would be regarded as truly astonishing today. On the basis of his analysis of the Army Alpha data that he collected during World War I on recruits of different nationalities (1923), Brigham inferred that members of some nationalities were inherently less intelligent than those of other nationalities. The implication of his conclusion was that since the United States was then accepting a large number of immigrants from the “less intelligent” nations and was also allowing “miscegenation” between white and African Americans within its borders, it would soon suffer a serious and dangerous decline in the general intellectual level of its population.

Little doubt exists that Brigham’s writings contributed to the repressive social climate in the United States in the 1920s. There is no evidence to show that his work had a direct influence on Congress in its formal deliberations on whether to restrict foreign immigration selectively. However, it would not be unreasonable to assume that his wealthy friends and mentors, all dedicated eugenicists who were well acquainted with his work—Madison Grant, Charles W. Gould, and Charles B. Davenport, among others—may have exerted considerable influence on Congress, which passed the restrictive immigration laws of 1924.

Some time after the publication of his 1923 book, Brigham had a change of heart. Exactly what caused this change is hard to say. In part, it may be attributed to some of T. L. Kelley’s writings (e.g., *The Influence of Nurture upon Native Differences* [1926] and *Crossroads in the Mind of Man* [1928]), in which Kelley enunciated his position that intelligence was not a simple unitary trait but separable into at least two component abilities—verbal and mathematical. Whatever the cause, in 1928, at a meeting of eugenics advocates, Brigham withdrew support from his earlier claims. In 1930, while secretary of the American Psychological Association, he published a formal retraction of his earlier

views in an article entitled, “Intelligence Tests of Immigrant Groups,” characterizing the conclusions of his earlier book as “pretentious” and “without foundation.” While it is perhaps likely that Brigham may have retained his own private racial and ethnic prejudices, his scientific writings showed no support for them after the middle 1920s. Brigham’s courage and dogged determination to fulfill his mission as an objective scientist mark him as a man of unusual distinction.

Perhaps most important among Brigham’s substantive contributions was his determination—possibly influenced, as suggested above, by T. L. Kelley’s writings—that intelligence, or academic aptitude, could not be adequately expressed in one overall score, as implied by the IQ. Rather, it required at least two separate and distinct evaluations of the talents of individuals, and he decided to report SAT results in two scores rather than one—describing separately the students’ verbal and mathematical abilities. This change of philosophy was a significant departure from the prevailing British view of intelligence as a single trait and, also as suggested above, it was at least partly responsible for his repudiation in 1930 of the Army research on which he had reported in 1923.

Brigham and his statistician, Cecil Brolyer, who had previously worked on mental tests with both Thorndike and Terman, first introduced the use of the three-digit SAT scale, defined with a range from 200 to 800, a mean of 500, and a standard deviation of 100, for reporting scores on the SAT. The value of the three-digit scale was that it avoided confusion with the then current 50–10 scale. The new three-digit scale also differed from the percentage mastery scale, then (and even now) in popular use in the schools and applied at that time by the College Board itself in reporting scores on the objectively scored Achievement Tests, which were introduced for regular use in the early 1940s. As a matter of routine, Brigham also introduced and performed rigorous item analyses to an extent unprecedented in testing practice, to examine the difficulty and discriminating power of test items being considered for operational use. (In 1932 he published *A Study of Error*, a volume summarizing and evaluating in detail a number of item types under consideration during the early years of the SAT and presenting the results of the item analyses conducted in the study of those items.) He also introduced the highly effective

technique of administering a nonoperational section (on which responses did not count toward the examinee's score) in regular operational forms of the test, in which untried items were pretested and new items types tried out on regular examinees. This practical, and highly innovative, device afforded him the opportunity to conduct experimental tryouts on a rigorously designed sample of regular examinees without the concern that they might be undermotivated and therefore perform more poorly than they otherwise would. It is noteworthy that these techniques of item tryout and item analysis, having proven their worth over the years, are even now in regular use by Educational Testing Service, some sixty-five years after they were first introduced by Brigham.

At about the same time, Brigham and Brolyer introduced a new index of item difficulty, the "delta." Although the delta is simply a normal transformation of the percentage-pass index of item difficulty and, like the percentage-pass index, is still sample dependent, the delta was a decided improvement because it tended to equalize the separations between successive values on the scale. The difference in difficulty between percentage-pass figures of .90 and .95 are much greater than between .50 and .55; corresponding differences on the delta scale are much more nearly equal and therefore much less likely to lead to interpretive error.

Brigham also introduced a regular program of validity studies, in which he was able to monitor the predictive power of the operational SAT and of proposed new item types in the individual colleges. This devotion to empirical checks remains a tradition at Educational Testing Service today and is implemented in the modern Validity Study Service of the College Board, a service offered without fee to all member colleges of the College Board.

In most ways, Brigham's views of the values and limitations of tests are as valid today as they were sixty-five years ago when he was actively engaged in developing the SAT. He urged that test scores be interpreted only in the context of other information, never alone, and he counseled against an overdependency on test scores. He also suggested that test scores should be used, along with other data, in counseling and guiding students in making educational choices and in helping the colleges make placement decisions

for students. Tests, he maintained, are not only useful for assessing present status and for making the usual predictions, but also for studying the kinds of errors that students typically make in their coursework and using that information to adjust and revise educational programs appropriately.

Thus, Brigham's contributions are more appropriately described as of an applied nature than as newly broken ground in theoretical development. His insistent adherence to rigorous psychometric principles and his search for new scientific procedures in the context of an ongoing testing program marked the beginnings of a major, larger-scale, nationwide, scientific testing effort that has remained the model of psychometric quality throughout the world.

(See also: SCHOOLING.)

BIBLIOGRAPHY

- BRIGHAM, C. C. (1923). *A study of American intelligence*. Princeton, NJ: Princeton University Press.
- BRIGHAM, C. C. (1930). Intelligence tests of immigrant groups. *Psychological Review*, 37, 158–165.
- BRIGHAM, C. C. (1932). *A study of error*. New York: College Entrance Examination Board.
- DOWNEY, M. T. (1961). *Carl Campbell Brigham*. Princeton, NJ: Educational Testing Service.
- KELLEY, T. L. (1926). *The influence of nature upon native differences*. New York: Macmillan.
- FRANKLIN, J. H. (1965). *From slavery to freedom: A history of American Negroes*. New York: Knopf.
- VALENTINE, J. A. (1987). *The College Board and the school curriculum*. New York: College Entrance Examination Board.

WILLIAM H. ANGOFF

BROCA, PAUL (1824–1880) Pierre Paul Broca was born on June 29, 1824, in Sainte-Foy-La-Grande, France, and died July 8, 1880, in Paris. A physician and surgeon during the critical transition between the prescientific and scientific periods of neurobehavioral science, Broca became a pivotal, yet controversial, figure in the history of this knowledge domain. The son of a country physician, Broca entered medicine with the expectation of assisting his father in

a rural family practice. His intellectual strength and academic ambitions, however, led to extensive cross-disciplinary contributions, spurred perhaps by the intensely competitive Parisian scientific and academic climate. Through an early association with François Lauret, psychiatrist and anatomist at Bicêtre in Paris, Broca was exposed to the revolution in the treatment of major psychiatric disorders and to the study of brain anatomy. This ultimately led to his becoming the youngest aide ever to be appointed in the Société Anatomique, where he pursued studies on cancer, surgical anesthesia produced by hypnotism, and aneurysms.

Simultaneously, Broca developed an interest in anthropology, consistent with contemporary interests in evolution and archaeology as related to the antiquity of humans. The interface of his interests and activities with early discussions of intelligence became focused in the Société d'Anthropologie, which he established. Thus, Broca is identified with early controversies related to intelligence and the physical dimensions of the cranium as well as intelligence and race. The combination of dissatisfaction with craniometric methods and the influence of the book, *On the Comparative Anatomy of the Nervous System as Seen in its Relations to Intelligence*, by François Lauret and Pierre Gratiolet, prompted Broca to explore the localization of cortical functioning by examining the brains of patients with previously observed disturbances of central language function. Although the nineteenth-century phrenologists believed that the anterior lobes of the brain were “the organ of articulate speech,” they had not provided evidence of an anatomical lesion in those regions to account for language deficits. Broca’s description of lesions in the left posterior portion of the frontal lobe in one patient, and a more focal lesion in the second and third frontal gyri (brain convolutions) of the left hemisphere in another patient, provided the first documented evidence of an anatomical lesion underlying a nonfluent aphasia in both patients. Broca referred to the language disturbance as an “aphemia,” to distinguish this disorder from a more general “loss of memory of words.” He subsequently presented additional cases of aphemia in which the lesion was located in the third frontal convolution. Seemingly contradictory cases were introduced by Charcot, one a case of

aphemia with lesions in the left parietal area and another a case in which a right frontal lobe lesion did not produce aphemia, were instrumental in convincing Broca of the left-hemisphere dominance for speech.

Accordingly, Broca is credited with having documented the anatomical basis of nonfluent aphasia, now also referred to as Broca’s aphasia. An annotated bibliography in support of his view was included with his application for membership in the Académie de Médecine and presented in the correspondence column of the academy’s bulletin. In the same issue, a letter submitted by Dr. Gustav Dax referred to a posthumous memoir of his father that provided evidence that Dax’s father, Marc, had much earlier presented details of forty patients with left-hemisphere lesions who evidenced speech impairment. Postmortem pathological confirmation of the lesion was not obtained by Dax, but there was indication in all cases of a right hemiplegia, thus pointing toward a basis in left-hemisphere pathology for the speech disturbances. Gustav Dax contended that his father had addressed the Southern Medical Congress in Montpellier, France, in 1836, and published a copy of his father’s work in 1865. Literature searches by Broca and by Dr. Gordon, librarian of the faculty of the University of Montpellier, failed to reveal any reference to the presentation to the Southern Medical Congress or to an earlier publication. Broca then published his most definitive paper on the subject of aphemia, in which he considered the Marc Dax memoir and also discussed the localization of aphasic disorders in the left hemisphere as well as the relationship between lateral dominance for speech and hand preference.

R. J. Joynt (1964) cited attempts by Pierre Marie, a well-known contemporary neurologist, to challenge Broca’s anatomical and clinical observations. Marie contended that Broca’s findings were superficial and ignored involvement of more posterior portions of the temporal lobes and the presence of diffuse neuropathological changes across these two cases. Broca’s biographer, Francis Schiller, in turn, examined the brains of these patients and challenged the assumptions and conclusions of Marie.

Although Broca is known principally for his role in the localization of expressive language function in the left prefrontal region, he has also been cited by influ-

ential, more contemporary neuroanatomists, such as Ramón y Cajal, who acknowledged Broca's contribution to the study of the anatomy of the limbic system. The latter led him to conclude that the limbic lobes were not exclusively involved with olfaction, a conclusion that was later proposed by two scientists who are today more widely known to have pioneered the anatomy and physiology of the limbic system, J. W. Papez and P. D. McLean. Others before Broca had described the relatively greater size of the frontal lobe in higher mammalian species, but Broca called attention to the proportional decrease in limbic size relative to the frontal lobes as a function of the place in the ascending order of species within the mammalian class.

Broca was known as an uncharacteristically political physician/scientist; he was elected to the French senate despite opposition from the more prevailing conservative and religious forces of the period. He died prematurely only five months after the election, presumably from a heart attack. Active during the very uneasy period of a shift in emphasis from a philosophical rationalism to a scientific empiricism during the mid- to late nineteenth century, Broca was noteworthy primarily for his contributions to the neurology of language and the anthropological interests that helped establish an early basis for the study of individual differences in intelligence as pursued subsequently by William Stern, Alfred BINET, and many others.

(See also: BRAIN; BRAIN, PATHOLOGIES OF THE.)

BIBLIOGRAPHY

- BENTON, A. L. (1984). Hemispheric dominance before Broca. *Neuropsychologia*, 22, 807–811.
- BROCA, P. (1856). *Des anéurysmes et de leur traitement*. Paris: Lubé et Asselin.
- GOLDSTEIN, K. (1953). Paul Broca. In W. Haymaker & F. Schiller (Eds.), *The founders of neurology* (pp. 12–15). Springfield, IL: Charles C. Thomas.
- JOYNT, R. J. (1964). Paul Pierre Broca: His contribution to the knowledge of aphasia. *Cortex*, 1, 206–213.
- JOYNT, R. J., & BENTON, A. L. (1964). The memoir of Marc Dax on aphasia. *Neurology*, 14, 851–854.

SCHILLER, F. S. (1979). *Paul Broca, founder of French anthropology, explorer of the human brain*. Berkeley: University of California Press.

MANFRED J. MEIER
DAVID KENT MERCER

BURT, CYRIL L. (1883–1971) Sir Cyril Burt was the twentieth century's most dramatic figure in British psychology. His research on intelligence and his strong personality generated both admiration and controversy. His legacy is one of both awe and anger in scientific circles. His posthumous reputation has been attacked and defended repeatedly in three major books and in numerous reviews and commentaries. Those who knew Cyril Burt personally concur that he was not an easy man to like. Sometimes arrogant, imperious, and always opinionated, Burt made enemies in high places and kept most supporters at arm's length.

During his extraordinarily long career, Burt founded educational psychology in Britain, contributed to the development of FACTOR ANALYSIS, developed the hierarchical model of intelligence, explored nature–nurture issues with studies of social-class differences in intelligence and school achievements, used his study of twins reared apart to defend his strong genetic view of intelligence, and championed educational selection by intellectual merit. Burt's research and opinions were widely published in the journal he founded, the *British Journal of Educational Psychology*, and in other scientific journals. His impact on educational policy in the United Kingdom is widely acknowledged by supporters and critics alike. He was the first psychologist knighted for his contributions to Britain.

After graduating from Oxford in 1906 and serving as lecturer in experimental psychology at the University of Liverpool, he was appointed chief psychologist for the London County Council in 1912, a post he held until 1932. During his twenty years as a school psychologist, Burt developed individual IQ tests, based on Alfred BINET's and Théophile Simon's work in France, and group intelligence tests that could be administered in the schools. His practical tests led to the reliable assessment of mentally retarded children and to the educational selection system in the United Kingdom, which is still partially in force.

In 1932, he was appointed to the most prestigious psychology chair in the United Kingdom, Charles SPEARMAN's chair at University College, London. Although he retired from that chair in 1950 (at the age of 67), he continued to write voluminously on nature–nurture issues, intelligence, and school achievements until his death in 1971.

Because Sir Cyril Burt was an outspoken advocate of genetic influences on variation in intelligence and achievements, he attracted many critics as well as followers. In 1974, Arthur JENSEN (a supporter), Ann and Alan Clarke (former students), and Leon Kamin (a critic) simultaneously published critiques of Burt's data on twins, raising questions about their veracity. All reached the conclusion that Burt's successively published figures for the twins did not add up.

In 1976, his critics, led by Oliver Gillie of the London *Sunday Times*, the Clarkes of Hull University, and other British psychologists of liberal/radical political persuasions, such as Liam Hudson, accused Cyril Burt publicly of scientific fraud. Specifically, they accused him of fabricating data on identical twins reared apart, a subject about which he had published several reports showing high correlations for general intelligence and other cognitive abilities. Based on Jensen's and Kamin's analyses of published reports, critics concluded that Burt had invented the separated twins completely. In addition, critics charged that Burt had invented assistants, collaborators, and co-authors, who he said collected twin data from the 1930s to the 1950s.

Three biographies were written about Cyril Burt, one critical (Hearnshaw, 1979) and two exonerating (Joynson, 1989; Fletcher, 1990). Hearnshaw accused Burt not only of fabricating data on twins but also of exaggerating his contributions to the development of factor analysis and of distorting his findings on social class differences in intelligence. From 1979 to 1989, Hearnshaw's biography seemed to seal Burt's fate as a fraudulent misanthrope. The British Psychological Society, which had previously held him in high regard, officially condemned him as a fraud.

Hearnshaw claimed to have reached his negative conclusions about Burt after careful examination of Burt's diary, extensive correspondence, and voluminous records left at University College, London, and in the University of Liverpool archives. Later, Burt's defenders claimed that Hearnshaw's examina-

tion of Burt's career was anything but complete or impartial.

Opponents of Burt's views of intelligence as a function of heredity and of social class differences dominated Hearnshaw's thesis that Cyril Burt was a mentally unstable fraud in his later years. Not only were his data on identical twins reared apart unreliable, perhaps nonexistent, but his earlier research on social-class differences in tested intelligence and his claims to major contributions to the development of factor analysis were unjustified. In 1979, Burt's defenders were confused and demoralized by what appeared to be scholarly evidence presented by Hearnshaw. Ten years later, Joynson and Fletcher undid Hearnshaw's reputation and revived Cyril Burt's.

Oddly, this nasty confrontation over intelligence and achievements had no implications for psychological studies of intelligence. Cyril Burt's research results (alleged or real) and his conclusions about individual and social-class variation in intelligence were entirely in agreement with the preponderance of contemporary research.

THREE DISPUTED ISSUES

Hearnshaw's attack and the Joynson/Fletcher defense of Cyril Burt's work hinged on three disputed issues:

1. Cyril Burt claimed that his own contributions to factor analysis (a method of extracting a few dimensions from diverse tests) were inspired by Karl Pearson's 1904 lecture at Oxford, where Burt was an undergraduate. By giving credit to Pearson, Burt seemed to diminish Charles Spearman's claim to be the father of factor analysis. Spearman supporters were enraged with Burt's claim.

Unless one was devoted to the method of factor analysis, this debate did not seem important. To measurement experts, Burt was clearly correct (e.g., Cronbach, 1979; Jensen, 1990). Spearman and Burt both contributed significantly to the development of factor analysis, and Burt's acknowledgment of Pearson's ideas was hardly reprehensible.

2. Cyril Burt claimed to have collected massive data on the IQ scores of London school children and their parents during the fifty years he was employed

as a school psychologist by the London borough and as a professor at the University of London. Critics doubted that he analyzed any data; rather, they proposed that he published his own prejudices about IQ and social class.

That Burt had access to large bodies of ability test data on children by social class cannot be doubted. He developed many of the tests and oversaw their administration, scoring, and interpretation. Because Burt's data were destroyed immediately after his death, it is impossible to know whether he actually had the data in his possession. He certainly could have visited files in the central school offices in the preceding fifty years to compile his reports. Were it not for allegations about other missing data, the school data would seem less suspicious.

Burt's interpretations of the social-class and IQ data are consonant with contemporary theory about transmission of intelligence from parents to children and the relationship between ability and social mobility in industrialized societies. Neither the reported results nor Burt's conclusions vary much from those of other investigators, whose data are not in doubt (e.g., Herrnstein; Horn, Loehlin & Willerman; Jensen; Scarr & Weinberg).

It is ironic that Burt is accused of class bias; in fact, he developed tests to select bright children from working-class families into elite grammar schools and universities. Tests are less class biased than teachers are, he thought. Burt was fond of pointing out that the majority of bright students came from disadvantaged backgrounds. The same class bias exists in the United States, based on ignorance of the benefits that standardized tests can afford talented but socially disadvantaged children.

3. Did Cyril Burt fudge, or even invent, data on identical twins reared apart to support his views on heredity? Burt published several reports from 1943 to 1960 on IQ test scores, educational attainments, and physical development of genetic relatives, including identical twins reared apart. As the samples reportedly grew, some of the correlation coefficients did not change—an unlikely outcome, which led both Arthur Jensen and Leon Kamin to question the validity of Burt's reports.

In Joynson's view, Burt probably reported the same correlations in subsequent articles when no new data had been collected. However, the unchanging correlations were not concentrated in the intelligence data, nor in the identical twin correlations. Rather, most of the unchanging correlations were in the physical data and for relatives other than separated twins.

The circumstances under which Burt collected data on separated twins are hazy. Exactly when, where, and how he came across separated twins and who actually tested them are matters of great dispute. Critics suspect that no such tests were ever given. Supporters take a more benign view: That Burt came across separated twins in his work with the schools and collected the pairs across many years. But where are the data? Given the emergency evacuation of the University of London to Wales during World War II and the bombing of records at University College, it is not surprising that Burt was separated from his data. He claimed to have found them serially over twenty-five years. No one will ever know whether Burt had separated-twin data or not.

Consensus exists among combatants in the Cyril Burt scandal on one point: He was careless, if not devious, about reporting details of samples and testing procedures. Carelessness may render his data useless but does not itself constitute proof of scientific fraud.

The loss of Cyril Burt's intelligence data is largely irrelevant to contemporary studies of intelligence. As even his critic Liam Hudson (1989) suggested, scientific standards have changed over the past century. More important certainly are his contributions to measurement, his development of ability tests, and his theoretical work on the nature of intelligence. Others have collected data that support Burt's major theoretical contributions.

(See also: HERITABILITY; HIERARCHICAL THEORIES OF INTELLIGENCE; TWIN STUDIES.)

BIBLIOGRAPHY

- CLARKE, A. M., & CLARKE, A. D. B. (EDS). (1974). *Mental Deficiency: The changing outlook* (3rd ed., pp. 168–171). London: Methuen.

- CRONBACH, L. J. (1979). Review of Hearnshaw (1979). *Science*, 206, 1392–1394.
- FLETCHER, R. (1990). *Science, ideology, and the media: The Cyril Burt scandal*. London and New York: Routledge.
- HEARNSHAW, L. S. (1979). *Cyril Burt: Psychologist*. London: Hodder & Stoughton.
- HUDSON, L. (1989, November 3–9). Recalling a scapegoat. *Psychology*, 1201–1202.
- JENSEN, A. R. (1990). Scientific fraud or false accusations? The case of Cyril Burt. In D. J. Miller, & M. Hersen (Eds.), *Research fraud in the behavioral and biomedical sciences*. New York: Wiley.
- JENSEN, A. R. (1974). Kinship correlations reported by Sir Cyril Burt. *Behavior Genetics*, 4, 1–28.
- JOYNSON, R. B. (1989). *The Burt affair*. London: Routledge.
- KAMIN, L. J. (1974). *The science and politics of IQ*. Hillsdale, NJ: Erlbaum.

SANDRA SCARR

C

CARROLL, JOHN B. (1916–) John B. Carroll is among the most prolific and influential of the world's psychologists and students of intelligence. His publications number 475 (and counting) and span more than fifty-five years, from his study of diversity of vocabulary usage (1938) to his spectacular review and synthesis of theory and research on human cognitive abilities (1993). A seminal contribution was *The Study of Language* (1953), which effectively introduced linguistic issues to psychology. His work is widely cited in today's literature on intelligence, psychometrics, reading instruction, psycholinguistics, and foreign language learning. In the literature surveyed for a recent annual Social Sciences Citation Index (Institute for Scientific Information, 1993), seventy-five of his writings (from 1945 to "in press") were cited at least once.

John Bissell Carroll was born June 5, 1916, in Hartford, Connecticut, of an English father who was in the insurance business and a New England mother and homemaker. Carroll (1980, 1985) reports, of his early years in Hartford, that before grade school he was taught to read by his aunt. Prior to high school, he read "history, antiquities, and the best American and English authors of the nineteenth century" in his grandfather's library. In Hartford Public High School, he learned not only Latin but also Greek—"three years of it"; in his spare time, he read texts in French and German and studied in the Hartford Library the grammars of Sanskrit and Armenian.

At age 13, Carroll attended a lecture at the Hartford Children's Museum on the Aztec and Mayan Indians of Mexico by Benjamin Whorf, who mentioned his interest in Nahuatl and other native Mexican languages. Whorf worked for a fire insurance company in Hartford and was a student at Yale with Edward Sapir. Carroll talked with Whorf after the lecture, and for many years they met regularly and worked together on translations from the Nahuatl and on Whorf's efforts to decipher Mayan hieroglyphics. Carroll also learned phonetics, phonemics, and other aspects of linguistics that Whorf was learning from Sapir.

Carroll majored in classics at Wesleyan University, and graduated with highest honors in 1937. In 1936, with Whorf's encouragement, Carroll had become a member of the Linguistic Society of America. Also at Whorf's suggestion, Carroll attended the Linguistic Institute at the University of Michigan in the summer of 1937, where Sapir was a faculty member. Whorf and Sapir discouraged Carroll from graduate study in linguistics, the discipline of his choice, because career opportunities appeared to be limited. He was encouraged to study the psychology of language instead. In 1937, at the University of Minnesota, he became the first graduate student of B. F. Skinner, whom Carroll (1980) describes as "a promising psychologist with an interest in language and language behavior."

Carroll's dissertation project was influenced less by Skinner's interests than by his own, which included

applying statistics to language (Carroll, 1938), and by those of L. L. THURSTONE, whose visiting lecture at Minnesota introduced Carroll to factor analysis as a means for deriving the structure of verbal abilities. During the summer of 1940, at Thurstone's laboratory at the University of Chicago, Carroll analyzed data he had collected that spring. A version of his dissertation, "A factor analysis of verbal abilities," was published in *Psychometrika* (Carroll, 1941) and continues to stand as a valid analysis of the structure of verbal abilities.

From 1940 to 1944, Carroll was successively instructor in psychology and education at Mount Holyoke College, instructor in psychology at Indiana University, and lecturer in psychology at the University of Chicago. His teaching assignments included courses in statistics, tests and measurements, and personnel psychology. As a personnel psychologist, he served from 1944 to 1946 as aviation psychologist, U.S. Naval Research (ensign to lieutenant), and from 1946 to 1949 as research psychologist, Department of the Army, in Washington, D.C.

In 1949, Carroll accepted a faculty position at the Harvard Graduate School of Education in educational measurement and statistics, where he rose from assistant professor to Roy E. Larsen Professor. From 1967 to 1974, he was senior research psychologist, Educational Testing Service, Princeton, New Jersey. In 1974, he became William R. Kenan, Jr., Professor of Psychology and director of the L. L. Thurstone Psychometric Laboratory at the University of North Carolina at Chapel Hill. In 1982, he formally retired from the faculty, but remained active in both research and teaching.

Carroll has been a member of the Linguistic Society of America since 1936 and a life member since 1952. A member of the American Psychological Association (APA) since 1939, he was elected a fellow in 1948 and president of the Division of Educational Psychology for 1966–1967. His membership in the Psychometric Society dates from 1940; he was also its secretary (1952–1955) and president (1960–1961). In 1965, he became a founding member of the National Academy of Education and served as vice-president from 1977 to 1981. He also has been a member of the American Association for the Advancement of Science (where he was a Fellow), the Modern Language Association of America, the Psychonomic Society, the National

Council of Teachers of English, the National Conference on Research in English, the Association for Computational Linguistics, and the American Council of Teachers of Foreign Languages.

In 1970, Carroll received the E. L. THORNDIKE Award from the Division of Educational Psychology of APA for distinguished service to educational psychology. In 1971, he was awarded the Diamond Jubilee Medal by the Institute of Linguistics (London), and in 1980 he won the Educational Testing Service Award for Distinguished Service to Measurement. The University of Minnesota bestowed on him the honorary Doctor of Science degree in 1986. He won the Distinguished Research Award from the National Conference on Research in English in 1990, and in that year was also honored at the annual Roundtable Meeting on Language Teaching by the School of Languages and Linguistics, Georgetown University. In 1993, he was invited to present the Spearman Lecture at the Spearman Seminar in Plymouth, England.

Among Carroll's significant publications directly related to intelligence are those of 1941, 1976, 1978, 1982, and 1993. The most extensive, by far, is *Human Cognitive Abilities* (Carroll, 1993). In that volume, Carroll reviews the history of intelligence measurement and evaluates competing theories of human intelligence. He reanalyzes a total of 461 earlier studies of the structure of cognitive abilities, employing a uniform method for analyzing these many sets of data. From the results, Carroll derives a "three-stratum theory" of cognitive ability, compares it with alternative theories, and presents its implications for public policy on ability testing and on other educational and social policy issues. The book is important not only for its textual content, but also for its comprehensive seventy-five-page list of references to the scientific literature on intelligence and its measurement.

(See also: FACTOR ANALYSIS.)

BIBLIOGRAPHY

- CARROLL, J. B. (1938). Diversity of vocabulary and the harmonic series law of word-frequency distributions. *Psychological Record*, 2, 379–386.

- CARROLL, J. B. (1941). A factor analysis of verbal abilities. *Psychometrika*, 6, 279–307.
- CARROLL, J. B. (1953). *The study of language: A survey of linguistics and related disciplines in America*. Cambridge, MA: Harvard University Press.
- CARROLL, J. B. (1976). Psychometric tests as cognitive tasks: A new “Structure of Intellect.” In Lauren Resnick (Ed.), *The nature of intelligence* (pp. 27–56). Hillsdale, NJ: Erlbaum.
- CARROLL, J. B. (1978). How shall we study individual differences in cognitive abilities?—Methodological and theoretical perspectives. *Intelligence*, 2, 87–115.
- CARROLL, J. B. (1980). The tale of a theoretical and applied linguistic psychologist. In Boyd H. Davis & Raymond O’Cain (Eds.), *Amsterdam studies in the theory and history of linguistic science III* (pp. 31–52). Amsterdam: John Benjamins B.V.
- CARROLL, J. B. (1982). The measurement of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 29–120). New York: Cambridge University Press.
- CARROLL, J. B. (1985). Author’s introduction. In Lorin W. Anderson (Ed.), *Perspectives on school learning: Selected writings of John B. Carroll* (pp. ix–xiv). Hillsdale, NJ: Erlbaum.
- CARROLL, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge: Cambridge University Press.
- INSTITUTE FOR SCIENTIFIC INFORMATION, INC. (1993). *Social sciences citation index, 1992 annual*. Philadelphia: Author.

LYLE V. JONES

CASE REPORTS The case report is a written record of a psychological evaluation. It is intended to communicate the results of the assessment in an easily understandable form that indicates what actions need to be taken. Reports take many forms, depending on the examiner’s orientation and training, the purposes of the evaluation, the background and qualifications of the targeted readers of the report, the complexity of the tests used in the evaluation, and so forth. Though they adhere to no single blueprint, virtually all case reports have common features, whether they are written for evaluations conducted in schools, clinics, hospitals, industries, prisons, or private practices; whether they focus primarily on tests of intelligence, personality, adaptive behavior, neuropsychological functioning,

educational achievement, or personnel selection; and whether they are intended for parents, teachers, psychologists, physicians, clients, or some combination of these audiences.

A case report should focus squarely on the individual being assessed, not on the tests used for that assessment. It must present clearly written, insightful interpretations of the test data that take fully into account the person’s particular background circumstances and test behaviors. It should avoid jargon and communicate in the language of the intended reader.

Case reports serve the dual functions of providing permanent records of evaluations and providing recommendations for future actions. Reports are organized in many ways, but the following sections should always be included: reason for referral, background information, appearance and test behaviors, test results and interpretation, summary, and recommendations.

Any evaluation is conducted for a reason, usually a problem of some sort. The reason for referring an individual for a psychological evaluation is stated clearly at the beginning of the report. The report should be written with the referral questions clearly in mind. Ideally, the report will include specific recommendations that address the referral questions.

The section on background information should include pertinent facts about the person’s medical history, educational history, family environment, socioeconomic status, interests and hobbies, previous test scores, and any additional information pertaining to the reason for referral. For the examiner’s protection and for future reference, one should document the sources of the information, for example, school records, a previous psychological report, medical records or reports, an interview with the person or a parent or therapist, and so forth.

The section on appearance and test behaviors, sometimes a neglected or hastily written part of the report, requires care and attention. A crisp description of the individual’s appearance and characteristic behaviors helps paint a concrete picture of the examinee and enables the reader to visualize the person functioning within the testing situation. A keen understanding of the behaviors may also help explain some of the fluctuations in the test profile. Behaviors need to be interpreted, not merely listed; a person’s knocking some test materials on the floor can reflect poor

tolerance of frustration for a specific failure, generalized anger or hostility, passive aggression, or clumsiness. The reader needs to know the examiner's interpretations of the behaviors and to understand consistencies and exceptions to the behavior patterns noted. At the same time, the examiner should support hypotheses about the person with specific behavioral referents. References to the individual as distractible, anxious, insecure, overconfident, poorly coordinated, or verbose are more meaningful if they are supported with observed behaviors. And although the examiner's observations are usually limited to interactions during the evaluation, the report should address behaviors attributed to the person by others, as stated in the background section. It should also generalize the observed behaviors to probable behaviors in the real world—which a one-on-one testing situation decidedly is not.

The section on test results and interpretation should include a list of all tests administered. Major test results should be given first, with test scores expressed numerically and with appropriate attention to errors of measurement. Scores should be translated to metrics that will communicate effectively to the reader (for example, percentile ranks are more readily understood than standard scores), and should include common-sense verbal explanations of the numerical results. Statistically significant discrepancies within profiles should be indicated, and peaks and valleys in a profile should be interpreted in the context of the background information and clinical observations of behavior reported in the previous sections of the report. If several instruments are administered, the results of each separate test might reasonably be discussed sequentially. The interpretation of results, however, should always be integrated across instruments. A finding on a test of intelligence should be cross-validated on tests of language or visual-motor coordination or personality. For example, if an adolescent is believed to display a moderate-to-severe visual-perceptual problem on the Wechsler Block Design subtest and on tests of visual-motor integration, errors in the interpretation of details in some Thematic Apperception Test pictures might relate directly to the perceptual problem, and not, say, to a thought disorder.

In the test results section, hypotheses about the person's overall functioning are specified and the evi-

dence of support (and rejection) of these hypotheses in test data, background information, and test behaviors is evaluated. The hypotheses may relate to theories that underlie the tests. They need to be psychologically meaningful. They need also to have a practical focus: They should be relevant to the problem for which the person was referred. They should indicate recommendations for intervention or other action. The results section should reach a dynamic integration of everything that is known about the person prior to the assessment and learned during the actual evaluation.

All case reports require a summary. Some individuals may need to abstract the essential findings, as quickly as possible, from a stack of case reports, and a well-written summary facilitates that task. The summary should provide a concise statement of the reason for referral, key background facts, behavioral characteristics, test results, hypotheses, and recommendations for taking appropriate action. Such actions may include suggestions for a type of therapy, educational intervention, enrollment in a school or program, change in medication, further testing, examination by a specific type of physician, selection of a particular job or training program, or modification of an existing environmental situation. These suggestions must address the referral questions and deal with issues learned during the evaluation process.

The case report thus is a well-written, cohesive document focused on the person being evaluated. It serves as a valuable file copy for future reference. It offers an insightful, integrated interpretation of the test and behavioral data, and provides meaningful suggestions for dealing with the referral issues. There is software for generating computerized case reports, but good report writing requires the experience and skills of one who has carefully studied the behavior of people.

BIBLIOGRAPHY

- EXNER, J. E., JR. (1986). *The Rorschach: A comprehensive system: Vol. 1. Basic foundations* (2nd ed.). New York: Wiley. Includes approach to writing case reports from the vantage point of Exner's interpretive system; see especially the chapters "The finished interpretation" and "Some assessment issues."
- EXNER, J. E., JR. (1991). *The Rorschach: A comprehensive system: Vol. 2. Interpretation* (2nd ed.). New York: Wiley. Includes

many illustrative case reports of patients with various disorders using Exner's approach to Rorschach interpretation.

- FRIEDMAN, A. F., WEBB, J. T., & LEWAK, R. (1989). *Psychological assessment with the MMPI*. Hillsdale, NJ: Erlbaum. The chapter "Interpretation and report writing" provides information on writing the behavioral and interpretive sections of reports that feature the MMPI.
- GREEN, R. L. (1991). *The MMPI-2/MMPI: An interpretive manual*. Boston: Allyn & Bacon. The chapter "Interpreting the MMPI-2 profile" offers tips for writing the behavioral and interpretive sections of MMPI-2 case reports.
- GREGORY, R. J. (1987). *Adult intellectual assessment*. Boston: Allyn & Bacon. Includes the chapter "Guide to report writing with examples of psychological reports."
- HYND, G. W. (Ed.). (1983). *The school psychologist*. Syracuse, NY: Syracuse University Press. Includes illustrative case reports using projective techniques in the chapter "Projective personality assessment techniques," and using neuropsychological instruments in the chapter "Brain-behavior relationships in children: Neuropsychological assessment in the schools."
- KAMPHAUS, R. W. (1993). *Clinical assessment of children's intelligence*. Boston: Allyn & Bacon. Includes a chapter on "Report writing and oral reporting," with illustrative case reports that focus on intelligence tests for children.
- KAUFMAN, A. S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley. Includes illustrations of clinical case reports for children, integrating the WISC-R with other instruments.
- KAUFMAN, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn & Bacon. Includes illustrations of integrated clinical and neuropsychological case reports for adolescents and adults.
- REITAN, R. M., & WOLFSON, D. (1992). *Neuropsychological evaluation of older children*. Tucson, AZ: Neuropsychology Press. Includes over 300 pages of neuropsychological case reports on 22 children and adolescents with a variety of disorders; features Halstead-Reitan and WISC-R.
- SATTLER, J. M. (1988). *Assessment of children* (3rd ed.). San Diego, CA: Jerome M. Sattler. Includes a chapter on "Report writing," plus many illustrative case reports that feature intelligence tests for children.
- SHELLENBERGER, S. (1982). Presentation and interpretation of psychological data in educational settings. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school*

psychology (pp. 51–81). New York: Wiley. Offers guidelines for writing communicative reports for school-related assessments; it does not appear in the second edition of the handbook.

ALAN S. KAUFMAN
NADEEN L. KAUFMAN

CATTELL, JAMES McKEEN (1860–1944)

James McKeen Cattell was born May 25, 1860, the son of the president of Lafayette College, in Easton, Pennsylvania. His early childhood experiences were influenced by the intellectual environment of his home. Cattell attended Lafayette and received an A.B. in English before moving to Europe, where he studied philosophy with Rudolf Hermann Lotze in Göttingen. There he was introduced to a reconciliation between religion and positivism. Following Lotze's death, which affected Cattell profoundly, he briefly returned to the United States, where he studied philosophy at Johns Hopkins University. In 1883, he returned to Europe to continue his studies for three years with the preeminent psychologist of the time, Wilhelm Wundt. Cattell reported (1928) that he brazenly announced to Wundt, "Herr professor, you need an assistant, and I shall be your assistant." His purpose was to establish an objective approach to experimental psychological studies, and he managed to fashion some tools to do this.

After completing his work for his doctorate in 1886, Cattell went to Cambridge University and enrolled in St. James College, with the aim of studying medicine, as had many psychologists of that time (e.g., Wundt, Hugo Münsterberg, William James, and Sigmund Freud). He met and became interested in the work of Sir Francis GALTON, to whom Cattell said that he was the greatest man he had ever met (Cattell, 1929). Galton's work on individual differences—and its obvious importance to human intelligence—was of particular interest to Cattell. Galton's influence permeates much of Cattell's later work, especially the possibility of measuring all abilities and capacities and using statistical and experimental devices.

In 1889, Cattell accepted a position in psychology at the University of Pennsylvania, which is alleged to have been the first designated chair in psychology. In

1891, he accepted a post at Columbia University, where he worked until 1917, mainly as professor of psychology (but also as chair of philosophy and anthropology); during this time, Columbia produced more doctorates in psychology than any other university. Cattell counted Robert S. Woodworth and Edward Lee Thorndike among his students.

Cattell was 32 years old when the first convention of the American Psychological Association (APA) was held in Philadelphia. In the report by the secretary, Morris Jastrow, the following passage is noted: "The invitation of Prof. Cattell to meet at Columbia College, December 27 and 28, was accepted." At age 35, Cattell was the fourth president of the APA (1895), following G. Stanley Hall, George Trumbull Ladd, and William James. In 1901, Cattell was the first psychologist to be appointed to the National Academy of Sciences, although some have suggested that his election may have been attributed to his professional editorial activities as much as to his scientific work (see Hilgard, 1987). His election was an important historic event for psychology—since the academy was then recognizing the scientific value of psychology. Along with James Baldwin, Cattell founded the journals of the Psychological Review Company, including *Psychological Review*, which he coedited from 1894 to 1904. In 1895, he bought the journal *Science*, which, after Cattell's death in 1944, was acquired by the American Association for the Advancement of Science (AAAS).

Cattell's contributions in the field of intelligence were mainly (1) his effort to develop a standardized procedure for the psychological testing of college students, (2) his participation in the organization of various professional associations, and (3) his founding of organizations for applied psychology—such as the Psychological Corporation (which supplied psychological testing and services to business and industry)—and professional publishing. His early work with Livingston Farrand (Cattell & Farrand, 1896) on the physical and mental measurements of Columbia University students, was focused on the importance of standardized measurements of intelligence—a feature of intelligence testing that was soon to be provided by the early twentieth-century work of Lewis TERMAN and his Stanford-Binet Scale.

Cattell's work in mental testing was augmented by his enormous contributions to professional and science

publishing. He edited the *American Naturalist* (from 1917), *Popular Science Monthly* (from 1900), *Science Monthly* (from 1915), and *School and Society* (1915–1939). Throughout his professional career, he was interested in individual differences and, in 1906, he was responsible for the publication of *American Men of Science*—which gave brief biographies of 4,000 men and women. Cattell also evaluated the prominence of each scientist by awarding a star to eminent scholars; further, he introduced a system in which ten competent leaders in each scientific field evaluated the merit of respective scholars and assigned a rank order to each. (This technique seems to be fashioned after a similar rating scale of 200 eminent psychologists done by Cattell in 1903. At that time, William James was first, followed by Cattell.)

In 1917, Cattell was discharged from Columbia University after many years of contention with its president, Nicholas Murray Butler. After leaving, Cattell continued his editorial work, especially through his Science Press, and in 1921 he founded and became president of the Psychological Corporation. The corporation supplied business and industry with psychological services, including a number of tests in the field of mental measurements. In 1925, Cattell was elected president of the AAAS and, in 1929, president of the International Congress of Psychology. In 1932, he began publication of *Leaders in Education*.

James McKeen Cattell is today known as a pioneer in psychology, as a teacher, researcher, writer, and publisher—as one who established journals and organizations in the sciences, social sciences, and especially in psychology. His work on individual differences, likely inspired by Sir Francis Galton, and the establishment of standard procedures for the evaluation of mental abilities had direct impact on the development of his students and on modern measures of human intelligence.

BIBLIOGRAPHY

- CATTELL, J. M. (1890). Mental tests and measurements. *Mind*, 15, 373–381.
- CATTELL, J. M. (1906). A statistical study of American men of science, the selection of a group of one thousand scientific men. *Science*, 24, 658–742.

- CATTELL, J. M. (1923). The Psychological Corporation. *Annals of the American Political and Social Science Society*, 110, 165–171.
- CATTELL, J. M. (1928). Early psychological laboratories. *Science*, 67, 534–548.
- CATTELL, J. M. (1929). Psychology in America. *Science*, 70, 335–347.
- CATTELL, J. M., & FARRAND, L. (1896). Physical and mental measurements of the students of Columbia University. *Psychological Review*, 3, 618–648.
- HILGARD, E. R. (1987). *Psychology in America: A historical survey*. San Diego: Harcourt Brace Jovanovich.
- SOKAL, M. M. (1980). Science and James McKeen Cattell, 1894 to 1945. *Science*, 209, 43–52.
- SOKAL, M. M. (1984). James McKeen Cattell and American psychology in the 1920s. In J. Brozek (Ed.), *Explorations in the history of psychology in the United States*, Lewisburg, PA: Bucknell University Press.
- SOKAL, M. M. (Ed.). (1987). *Psychological testing and American society: 1890–1987*. New Brunswick, NJ: Rutgers University Press.

ROBERT L. SOLSO

CATTELL, R. B. (1905–) Raymond Bernard Cattell was born in Staffordshire, England, in 1905. Long a Distinguished Research Professor at the University of Illinois, he is regarded as one of the outstanding factor-analytic researchers in the fields of ability and personality. He was educated at Torquay Grammar School in Devonshire and won a county scholarship to Kings College, University of London. He studied chemistry there, graduating with a first class degree in 1924. He became interested in psychology, and completed his Ph.D. under Cyril BURT, while an assistant at the London Child Guidance Clinic. He was later awarded the DSc by London University for his contribution to the study of ability and personality.

After receiving his Ph.D. in 1927, he became a lecturer in the Education Department of the University College of the South West, Exeter (now the University of Exeter). In 1930, he married Monica Rogers and they had one child, Hereward Seagrave. In 1932 he became director of the City of Leicester Child Guidance Clinic, one of the first child-guidance clinics in

the United Kingdom, where he remained until 1937, when he left for America to work with Edward THORNDIKE. From 1938 to 1941 he was the G. Stanley Hall Professor of Genetic Psychology at Clark University, and from 1941 to 1943 he was a lecturer at Harvard. He was then appointed to the psychology department of the University of Illinois at Urbana as Research Professor of Psychology. He remained there until his retirement, when he took up an Emeritus Professorship at the University of Hawaii.

It was at Illinois that most of his best-known work in human intelligence and abilities was completed, but it should be noted that, unlike many academic research psychologists, Cattell had practical experience in the child-guidance clinic, and this experience informs all his theory and research. Similarly, his training with Burt, who was also an applied psychologist, was influential in this respect.

Cattell has published about forty books and more than 400 papers. These publications illustrate with great clarity the contribution of Cattell to the study of intelligence. As this brief description of his education and career indicates, Cattell is part of the so-called London School of Psychometrists, founded by Charles SPEARMAN. Indeed, together with H. J. EYSENCK he is one of its last survivors. This background is evident throughout his work, which falls into two stages. Some of Cattell's earliest research was concerned with the measurement of intelligence and its factor-analytic structure (1940, 1943). He then turned to the study of personality, for which he is probably best known (1977). In the 1960s, however, he turned his attention again to intelligence (1971), because improved factor-analytic methods (for which he and his colleagues were partly responsible) and more powerful computers made possible the resolution of the factor structure of abilities. His 1971 book contains a detailed account of his views of and work in intelligence. Various aspects of the field are dealt with, as set out below.

The Factor Structure of Abilities. Cattell is probably best known for his factor-analytic work. He has argued that FACTOR ANALYSIS is an ideal approach to the complexities of real-world psychology. He has developed and applied the techniques to a huge range of psychology. In the field of ability, he has claimed that there are two factors that are supremely important—FLUID AND CRYSTALLIZED INTELLIGENCE (1940,

1943, 1971, 1976, 1977, 1986, 1991). These two factors are those actually measured by old-fashioned intelligence tests such as the Wechsler scales and Stanford-Binet. *Fluid ability* is a kind of reasoning ability hypothesized to be highly influenced by genetic constitution. *Crystallized ability* reflects how fluid ability, affected by many environmental influences over the course of development, is manifested in the set of skills valued in a culture. Thus, in the United States or in Great Britain, crystallized ability can be seen in academic achievement, in high earning capacity, and in holding responsible and important jobs. In other cultures, fluid ability may be invested in different skills. In this view, the older notion of *g*, or general ability, has been split into these two correlated factors.

Measurement of Intelligence. For Cattell, measurement is central to a scientific psychology. Since, as has been seen, intelligence is not a unitary faculty, Cattell's intelligence tests differ from the standard measures. For fluid ability, Cattell developed the Culture-Fair Intelligence Test (1940, 1971), which is intended to minimize the effects of the environment and to measure, as far as is possible with a pencil-and-paper test, innate reasoning ability. For crystallized ability, Cattell used his Comprehensive Ability Battery (1971, 1976, 1991), which measures the separate but correlated factors that comprise crystallized ability.

The Inheritance of Intelligence. This is, perhaps, the most controversial aspect of Cattell's work on intelligence. He and his coworkers developed a statistically highly sophisticated approach to the study of the HERITABILITY of all traits, not just intelligence (1982). Applying these methods in samples of people gathered from different societies around the world, Cattell concluded that approximately 65 percent of the individual-differences variance in fluid ability can be attributed to genetic factors, the rest being determined by environmental factors and by measurement error. It must be stressed that the reference here is to variability *between individuals in samples*. The figures do not refer to any particular individual. This figure will vary in different populations, depending on their genetic similarity and the variability in the opportunities in the environment. In India, for example, where there is enormous variability in environmental opportunities compared with England, the environment is likely to be more influential. Cattell's biometric analyses allow

the study of the effects of gene dominance, assortative mating, and a further analysis of environmental influences into the shared and unique experiences of individuals.

Application in Occupational Selection. Cattell has always been concerned with the practical application of his work. Intelligence and ability testing are parts of his approach to occupational selection. Relevant tests are given, together with tests of personality and motivation, and selection for a particular job is determined by appropriately weighting the scores—what Cattell refers to as *specification equations* (1976, 1986).

Application to Educational Psychology. Cattell first worked in a department of education and in educational psychology. In the tradition of Burt, he reasons that it is impossible to estimate how well a child is doing at school unless his or her abilities (both fluid and crystallized ability) are known. Furthermore, for selection, as was the case with occupations, Cattell has demonstrated that a high level of prediction (with correlations of about .6) of academic success is possible if ability, personality, and motivation tests are used (1963, 1977).

Theoretical Psychology. Cattell's studies of intelligence were only one aspect of his factor-analytic approach to human psychology, which included the study of personality, motivation, states (1977), and in principle, at least, situations. Essentially, Cattell formulated a psychometric model of personality in which any behavior could be specified and predicted by an individual's weighted scores on the main ability, personality, and dynamic factors, measures of which can be found in Cattell's work (see Cattell & Kline, 1977, for a summary). Thus his occupational and educational work are really only instances of his more general theoretical model.

BIBLIOGRAPHY

Cattell's own writings tend to be difficult, but some books written with collaborators are more accessible.

CATTELL, R. B. (1940). A culture-free intelligence test. *Journal of Educational Psychology*, 31, 161–179. The first description of the Culture-Free Intelligence Test, measuring fluid ability.

CATTELL, R. B. (1943). The measurement of adult intelli-

- gence. *Psychological Bulletin*, 40, 153–193. An important paper discussing the evidence for the concepts of fluid and crystallized ability.
- CATTELL, R. B. (1971). *Abilities: Their structure, growth, and action*. New York: Houghton Mifflin. A detailed account of Cattell's views on human abilities, and the main source for this aspect of his work (although later research is obviously not included).
- CATTELL, R. B. (1982). *The inheritance of personality and ability*. New York: Academic Press. The detailed account of Cattell's methods of investigating the influences of genetic and environmental factors on psychological traits.
- CATTELL, R. B., & BUTCHER, J. (1963). *The prediction of achievement and creativity*. Indianapolis, IN: Bobbs-Merrill. The authors describe research showing the contributions of ability, personality, and motivational traits to the prediction of educational success.
- CATTELL, R. B., & KLINE, P. (1977). *The scientific analysis of personality and motivation*. London: Academic Press. The most simple account of Cattell's whole system; it enables readers to see his work on intelligence in context, although it concentrates on the research into personality and motivation.
- CATTELL, R. B., & JOHNSON, R. C. (1986). *Functional psychological testing*. New York: Brunner/Mazel. In this book can be found chapters describing the use of Cattell's tests, including the use of ability and intelligence tests in applied settings.
- HAKSTIAN, A. R., & CATTELL, R. B. (1976). *The Comprehensive Ability Battery: CAB*. Champaign, IL: IPAT. The CAB is Cattell's test for measuring the most important human abilities.
- KLINE, P. (1991). *Intelligence: The psychometric view*. The simplest account available of the factor analysis of human abilities. It describes the factor-analytic method, summarizes the findings of Cattell, and compares them with other factor analytic accounts.

PAUL KLINE

CHILD AND ADOLESCENT INTELLIGENCE It is clear that there is no single concept that one can point to and say, "That's it. That's what psychologists mean by *intelligence*." When psychologists get down to the business of studying intelligence, it means something different to each of them. This is

nowhere more true than in the study of intellectual development, where research has varied greatly in its concentration on empirical questions about which abilities are present at which age levels versus theoretical questions about the ultimate nature of intelligence.

At the extremities of the life-span, the empirical orientation has predominated. This is because questions about intellectual competence or the lack of it are significant at the beginning and the end of life. As life begins, it is natural to wonder how long it will be before abilities such as basic numerical concepts emerge (see INFANCY; INFANT TESTS AS MEASURES OF EARLY COMPETENCE). As life ends, it is natural to wonder which abilities are sensitive to the ravages of aging and which are resistant (see AGING AND INTELLIGENCE).

Between these poles, the study of intellectual development has been much more theoretical. During child-to-young-adult development, in particular, research has revolved around a succession of high-altitude pictures of what adult intelligence is like and, therefore, of where intellectual development is going and what its purposes may be. Four metaphors have emerged in twentieth-century research: *associationism*, *logicism*, *computationism*, and *intuitionism*.

ASSOCIATIONISM

The purpose of any metaphor is to make some sense of dark and inexpressible subjects and to help researchers decide what to study. During the first half of the twentieth century, psychology as a whole was dominated by a metaphor that had been imported from British empiricist philosophy, *associationism*. Associationism's core thesis was that the mind is governed by what John Locke had termed "the association of ideas." According to this theses, what we call reasoning or intelligence reflects an underlying organization in which ideas become interconnected because they have occurred close together in time (association by contiguity), or they resemble each other in some noticeable way (association by similarity), or they differ in some noticeable way (association by contrast).

The vehicle for associationism in psychology was stimulus-response learning theory. Learning was treated as the premier intellectual process, especially learning in the sense of forming associations between specific

stimuli (e.g., the sound of a bell) and specific responses (e.g., salivation). In intellectual development, emphasis was placed on distinctions between different types or styles of learning that were thought to be indigenous to different age levels. The most influential of these distinctions, historically, was between a primitive learning style that predominates during infancy and early childhood and a more advanced style that predominates during later childhood and adolescence (e.g., Stevenson, 1972). The primitive style consisted of rote memorization of the surface features of stimuli, and it was variously called absolute learning, compound learning, nonmediational learning, and object learning. The advanced style involved going beyond the surface features of stimuli and bringing conceptual knowledge to bear, and it was variously called component learning, concept learning, mediational learning, and relational learning.

The two learning styles can be understood by considering a procedure that was used to diagnose them—a technique called *discrimination transfer*, first popularized by Kenneth Spence in the 1930s. The procedure consists of two phases, which are illustrated in

Figure 1. During the first (or learning) phase, children study a series of stimuli that vary along some familiar perceptual dimensions (shape and size in the example). They are told that some stimuli are winners and the rest are losers and that their task is to learn which are which. Then they receive a series of trials in which they are asked to guess the classification of each stimulus, and they receive feedback about the correctness of each response. Children could learn this task in either of two ways: (1) by memorizing the classifications of individual stimuli (large square = winner, small square = winner, large circle = loser, small circle = loser), or (2) by acquiring a rule that applies to all of the stimuli (squares are winners). The second (or transfer) phase determines which style was operative during the first phase. Once the task is learned, the winner-loser classifications are changed without warning, and children must now learn the new classifications. As can be seen in Figure 1, the change involves either staying with the previously relevant dimension (top-right stimuli) and reversing classifications along that dimension (the two circles are now the winners), or switching to the other dimension

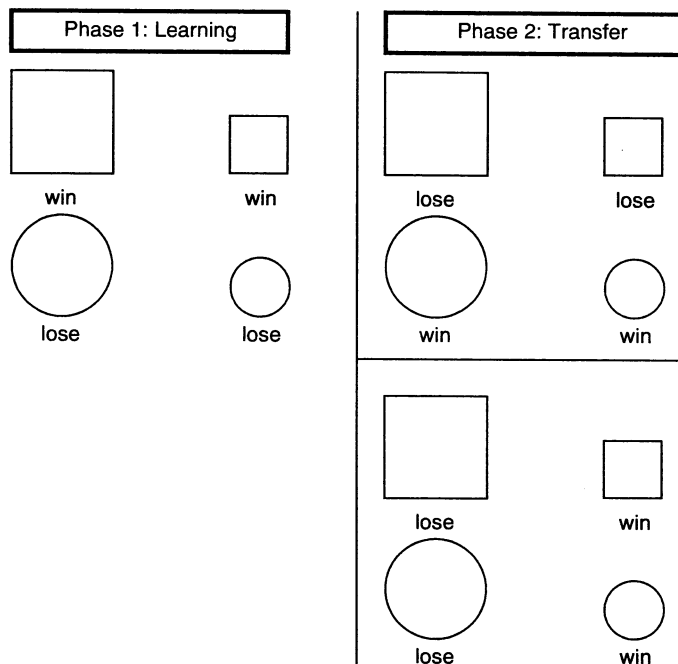


Figure 1

A typical discrimination transfer test for diagnosing the incidence of primitive and advanced learning styles in children

(bottom-right stimuli) and assigning classifications along that dimension (the small stimuli are now the winners). The second form of transfer will be easier if children simply memorized the original classifications of individual stimuli *because two of them retain their original classifications*. But, the first form of transfer will be easier if they learned a rule based on shape *because shape is still the relevant dimension*. Many studies conducted from the late 1940s onward showed that during childhood and adolescence, the primitive form of learning became progressively less frequent while the advanced form of learning became progressively more frequent (Kendler & Kendler, 1962).

LOGICISM

By the 1960s, it had become common to criticize discrimination transfer tasks on the ground that they were too simple to capture the complex changes in reasoning that are the real essence of intellectual development. The notion that a more complex, high-cognitive metaphor was needed crystallized after a conference on "Thought in the Young Child" that the Science Research Council sponsored in 1960. That conference focused on the work of the Swiss psychologist Jean PIAGET.

Piaget had promulgated a high-cognitive metaphor that went on to dominate research on child and adolescent intelligence for the next two decades. During

the 1930s, Piaget had studied mathematical philosophy intensively to prepare for a course in the history of science that he was teaching at the University of Geneva. As he later told the story, those studies convinced him that formal logic—the systematic generation of valid inferences through rigorous deductive analysis of premises—was an ideal metaphor for adult intelligence.

The adoption of this metaphor implied that intellectual development could be viewed as progress toward the sort of formalized logical reasoning operations that are found in mathematics and science (Piaget, 1953). Piaget implemented this view in many ingenious studies of the growth of mathematical and scientific reasoning during childhood and adolescence. Eventually, a grand, four-stage model of intellectual development emerged from this work (Piaget & Inhelder, 1969). Except for the first stage, which dealt with infancy, all of the stages were concerned with the acquisition of competencies that have a strong logical flavor (see PIAGETIAN THEORY OF DEVELOPMENT).

An example of a tool for studying transitive inference is exhibited in Figure 2. Children are shown a series of three rods (A, B, C) that differ in length by small amounts. The relationships between the adjacent pairs in the series are demonstrated for them ($A > B$, $B > C$). They are then asked to deduce the relationship between the nonadjacent pair ($A > C$) from these premises but in the presence of a visual illusion

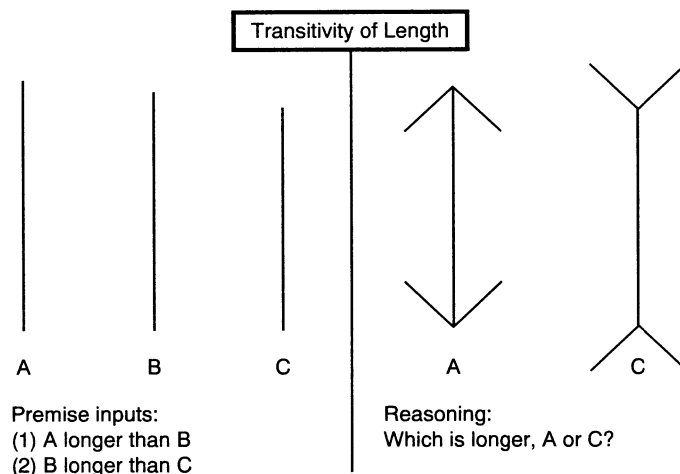


Figure 2
Piaget's transitive inference problem

(Muller-Lyer arrowheads) that makes it look as though the relationship is the opposite of what it actually is ($A < C$). Piaget found that preschool children (pre-operational stage) were illogical; they could not deduce the true relationship at all and were taken in by the visual illusion. Elementary school children (concrete-operational) exhibited some logical competence in that they could deduce the true relationship despite the visual illusion. However, they were not fully competent. Although they could deduce the relationship if premises were presented concretely (the rods in Figure 2), they could not if the presentation was purely verbal (e.g., John is taller than Jim. Jim is taller than Don. Who's tallest?). This more abstract logical competence was not detected until adolescence (formal-operational stage).

COMPUTATIONISM

In the late 1970s and early 1980s, logicism was supplanted by the third metaphor, computationism. Computationism retained Piaget's emphasis on studying the development of complex reasoning, but simpler enabling mechanisms (e.g., short-term memory capacity) also became important. It substituted the computer for deductive logic as the basal metaphor: Information was said to be "stored" in symbolic mental representations, and thinking was said to involve "retrieval" of computational operations followed by "computations" performed on information stored in the representations.

In the study of child and adolescent intelligence, the vehicle for this metaphor has been information-processing theory (e.g., Klahr & Wallace, 1976). The

key feature of this approach is that it views adult intelligence in terms of cognitive architectures called *information-processing systems*. These systems consist of (1) distinct memory stores for representing information; (2) operations for transferring information from one store to another; and (3) operations for performing computations on information housed in memory stores. A very simple example of such a system is shown in Figure 3. There are two memory stores, short-term (very small capacity) and long-term (very large capacity), two transferral operations (search and retrieve), and an indefinite number of computational operations stored in long-term memory. To solve a reasoning problem such as transitive inference (Figure 2), children first encode symbolic representations of the premise relationships into short-term memory. Then, they use the representations as cues to search and retrieve computational operations from long-term memory. Finally, they deduce the A-C relationship by computing the answer from the premise representations in short-term memory.

Computationism led to two critical advances in our understanding of intellectual development. First, because information-processing systems consist of distinct components—memory stores, transferral operations, computational operations—there were now more sources of developmental change than simply improvements in logical competence. As transitive inference ability improves, it could be, as Piaget thought, because basic logical competence is being acquired. It could also be because children are becoming better able to encode problem information into short-term memory, because the capacity of short-term memory

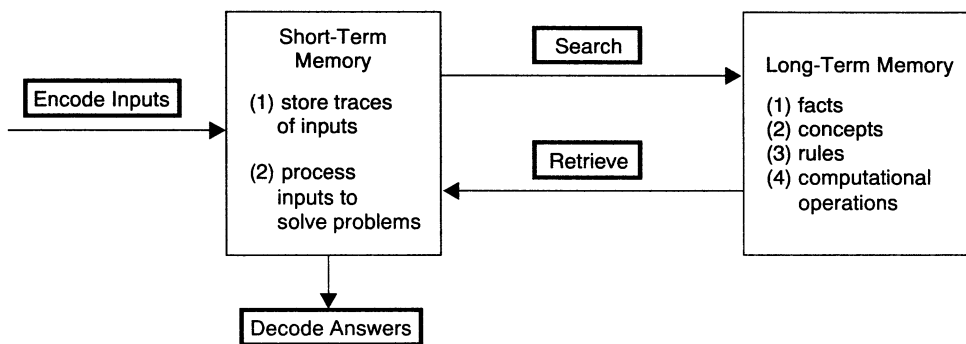


Figure 3

A simple information-processing system

is increasing, because children are becoming better at retrieving computational operations, and so forth. Second, a consistent finding of research has been that these noncompetence components play much more prominent roles in intellectual development than do improvements in logical competence. With many reasoning problems, it has been found that the basic logical competence to solve them is present in 4- and 5-year-olds, but that the information-processing components that are necessary to implement that competence develop gradually during childhood and adolescence.

INTUITIONISM

A fourth metaphor, intuitionism, has now become prevalent. Adoption of this metaphor was fomented by a shift in our understanding of just how logical adult intelligence is. The Piagetian and information-processing traditions agreed on a fundamental point: Logic is a good model for adult intelligence. The dispute between them was about whether age improvements in reasoning are due to acquisition of logical competence per se (Piaget) or to refinements in the support systems that implement that competence (information processing).

During the 1980s, the notion that adult intelligence is fundamentally logical was challenged by both developmental and adult research. In developmental research, it was found that some of the most elementary mathematical and scientific principles, ones that Piaget claimed were understood by late childhood, are never understood at any age level, even by mathematicians and scientists. Probability is a prominent example: Even statisticians regularly give answers to probability problems that violate the basic laws of probability theory. Other principles seem to be understood during childhood when one task is used to measure them, but seem to be absent even in college students when a slightly different task is used. For example, most third graders seem to understand conservation of weight if concrete objects are used: When asked if the relative weight of two clay balls will change if one of them is flattened into a pancake, most third graders will say no. But if college students are asked "Would you weigh more if you were standing or sitting," they almost never say "no difference." Turning to adult re-

search, studies of judgment and decision making in everyday situations (purchasing groceries, selecting medical treatments, choosing investment plans) have repeatedly found that even the simplest decisions do not involve logical reasoning, but instead, are governed by rules of thumb, biases, and other nonlogical principles.

Naturally enough, such results have encouraged the view that adult intelligence is overwhelmingly intuitive rather than logical. Fuzzy-trace theory (Brainerd & Reyna, 1990, 1993) has implemented this view in the study of intellectual development. This theory differs radically from the Piagetian and information-processing traditions in the sense that its objective is to explain how intelligence becomes progressively more intuitive during childhood and adolescence, rather than how it becomes progressively more logical. According to fuzzy-trace theory, intelligence is intuitive because of (a) gist extraction combined with (b) fuzzy-processing preferences. Point *a* refers to the fact that when problem information is encoded into memory, it leads to simultaneous retrieval of other memories (gists) that capture simple patterns in the information. Point *b* refers to the fact that there is a strong preference during reasoning for processing streamlined gist memories rather than memories of the actual problem information. In transitive inference, for example, it has been found that presentation of the adjacent relationships leads to retrieval of global spatial patterns such as "big things start on the left" (point *a*). It has also been found that inferences are made by processing that pattern with a simple perceptual operation ("A is to the left of C") rather than by applying logical operations to the actual premises (point *b*). Thus, in fuzzy-trace theory, interest centers on questions about how developmental changes in gist retrieval and fuzzy-processing preferences cause intelligence to become more intuitive, not on questions about how logical competence is acquired or implemented.

CONCLUSIONS

During the twentieth century, the progression of metaphors for adult intelligence has led to two shifts in the ways that we study child and adolescent intelligence. The first shift, which occurred during the 1960s, consisted of moving away from an emphasis on

age changes in simple learning styles toward an emphasis on age changes in complex reasoning, particularly mathematical and scientific reasoning. The second shift, which took place in the 1980s, is concerned with the psychological bases for such reasoning. Since the 1960s, it was thought that psychological versions of logical operations were responsible for reasoning. More recent research has favored the view that reasoning, whether simple or complex, is controlled by processes that are intuitive rather than logical.

(See also: CHILDREN'S CONSERVATION CONCEPTS; DEVELOPMENT, COGNITIVE; KAUFMAN ASSESSMENT BATTERY FOR CHILDREN; MENTAL AGE.)

BIBLIOGRAPHY

- BRAINERD, C. J., & REYNA, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, 10, 3–47.
- BRAINERD, C. J., & REYNA, V. F. (1993). Memory independence and memory interference in cognitive development. *Psychological Review*, 100, 42–67.
- KENDLER, H. H., & KENDLER, T. S. (1962). Vertical and horizontal processes in problem solving. *Psychological Review*, 69, 1–16.
- KLAHR, D., & WALLACE, J. G. (1976). *Cognitive development: An information processing view*. Hillsdale, NJ: Erlbaum.
- PIAGET, J. (1953). *Logic and psychology*. Manchester: University of Manchester Press.
- PIAGET, J., & INHELDER, B. (1969). *The psychology of the child*. New York: Basic Books.
- STEVENSON, H. W. (1972). *Children's learning*. New York: Appleton.

CHARLES J. BRAINERD

CHILDREN'S CONSERVATION CONCEPTS

The Swiss psychologist Jean PIAGET once asked a colleague why he had become a mathematician. The mathematician replied with a charming anecdote: Many years before, as a child of 5, he had been playing with some pebbles. He happened to count them. He then spread the pebbles out, thinking that there would be more than before. When he recounted them, however, the number had not changed. The mathematician

performed further experiments of this sort, discovering each time that the number had not changed. This discovery so astonished him, he said, that he later took up the study of mathematics.

Piaget, for his part, was so fascinated by the mathematician's astonishment that he turned this anecdote into one of the most extensively studied phenomena of childhood intelligence, the conservation concepts. This article sketches some of the basic things that a half-century of research has revealed about children's conservation concepts, organizing the presentation around things that do and do not support major claims that Piaget made about them.

First, it is necessary to define what a conservation concept is, which can be done with the aid of Figure 1. In its most general sense, the term *conservation concept* refers to a child's understanding that quantitative relationships between objects remain invariant when spatial (i.e., nonquantitative) transformations are performed on the objects. Assessment of particular conservation concepts involves administering problems that conform to the following script: There are two objects, which are designated *S* ("standard object") and *V* ("variable object"), respectively. At the start of a problem, two things are true about *S* and *V*: (1) they are known to be equal on some quantitative dimension (length, weight, volume, area, etc.), and (2) they are visually identical. Next, object *V* is subjected to a spatial transformation ($V \rightarrow V'$) of some sort that destroys the second property (i.e., V' looks very different from *S*) without affecting the first property (i.e., $S = V'$) because nothing was added to, or subtracted from, *V* to produce V' . Finally, children are questioned to determine whether they understand that 1 was conserved when 2 was destroyed—hence, the term *conservation concept*.

The three most frequently studied conservation problems—number, quantity, and length—are shown in Figure 1. The initial state for each problem is shown at the top: two identical rows containing five chips apiece; two identical glasses containing equal amounts of liquid; two identical lines of equal length. The transformed $V \rightarrow V'$ state is shown at the bottom: the white row has been spread out so that it is longer than the black row; the contents of one of the glasses have been poured into a taller and narrower glass; inward-pointing arrowheads have been affixed to one of the lines,

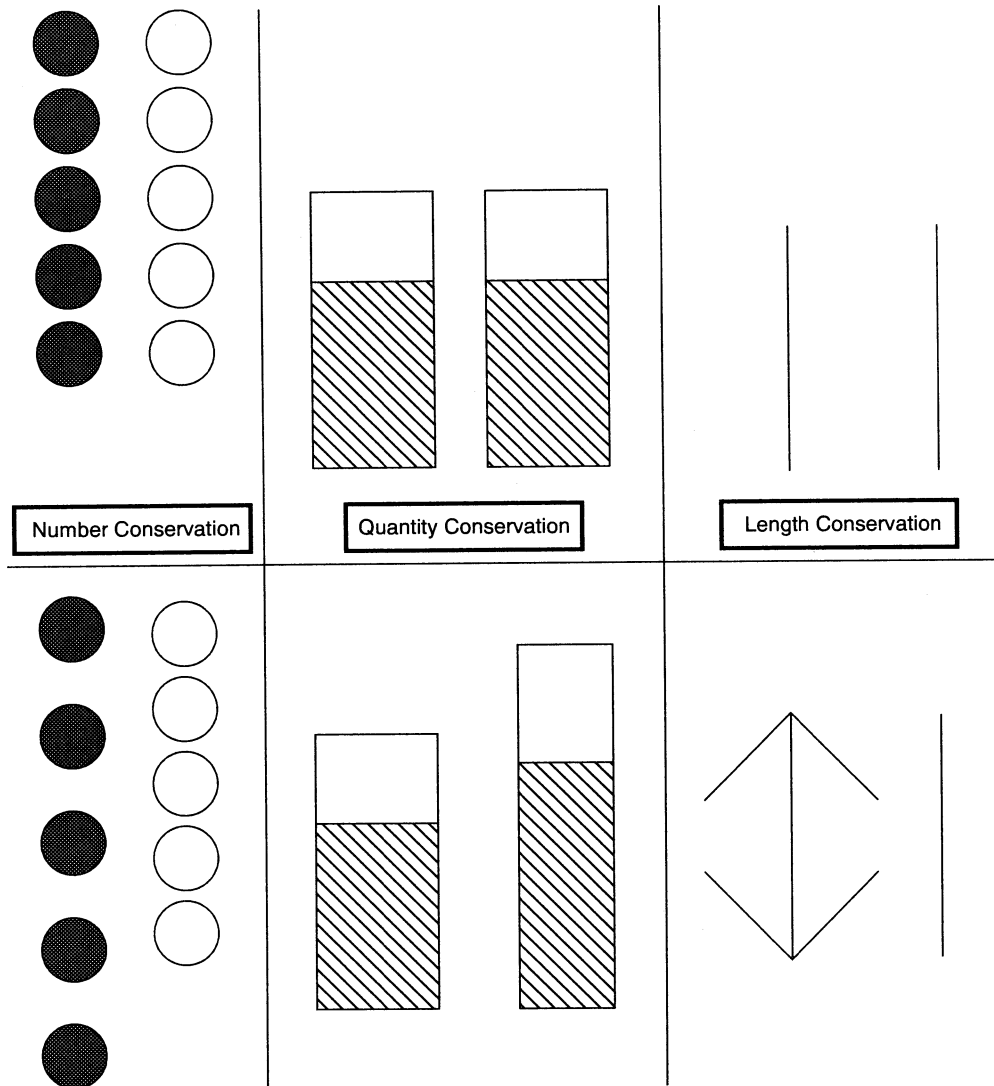


Figure 1
Examples of three conservation concepts

so that it looks shorter than the other line. Following such transformations, Piaget first asked children to make judgements about the quantitative relationships between S and V' : Is there still the same number of black chips as white chips? Do the two glasses still have the same amount to drink? Are the two lines still the same length? He then asked children to explain their judgments: How do you know that? A standardized test of these three conservation concepts, plus seven others (area, discontinuous quantity, distance, mass, two- and three-dimensional space, and weight), is available (Goldschmid & Bentler, 1968).

SUPPORTIVE EVIDENCE

During the 1930s and 1940s, Piaget conducted many studies of children's understanding of conservation; ultimately, this work became the centerpiece of his stage theory of cognitive development (e.g., Piaget & Inhelder, 1969). Without putting too fine a point on it, conservation concepts came to be regarded as the surest signs of adult-style reasoning; they were the crux of Piaget's definition of intelligence. Along the way, Piaget made a number of important claims about these concepts, some that were subsequently con-

firmed and others that were disconfirmed. Three examples of claims of the former sort are given here; nonsupportive evidence is presented in the next section.

Piaget's first, most basic proposal was that children simply fail to grasp these concepts before the mid-elementary-school years. It is impossible to recreate the skepticism with which this idea was greeted in the 1960s, when Piaget's theory was becoming influential in the United States. Some sense of it can be gleaned from the incredulous tone of questions that were common in writings of that era. How can children who have just counted the black and white chips, so that they know that the two rows have the same number, possibly believe that the longer row has more chips? How can children who have just poured 10 ounces of liquid into each glass, so that they know that the two glasses have the same amount, possibly believe that there is more to drink in the taller glass? Although it may seem utterly incongruous, they do. With problems like those in Figure 1, most children below the age of 6 fail them across the board.

Second, although all conservation concepts involve understanding a single overriding fact (namely, that visual appearance does not determine quantitative rela-

tionships), Piaget thought that children do not acquire all these concepts in unison. Instead, conservation development was characterized as a slow and laborious process, spanning all of childhood and the early part of adolescence. Piaget further maintained that individual concepts emerge in stable sequences, which he called "horizontal decalages." This claim, too, has received considerable support. Typically, several years intervene between a child's solutions of different conservation problems, and a good deal of agreement exists as to which ones are solved before others. With M. L. Goldschmid and P. M. Bentler's (1968) test, for instance, ten problems were originally standardized on samples of children between 5 and 9 years old. The proportions of children who gave a "still same" judgment on each problem are displayed in Figure 2. As can be seen, even though the mean age of these children was above 7, the average proportion of "still same" judgments across problems was only slightly above half (.56). There was also great variability in the "still same" rates for different problems. In Goldschmid and Bentler's (1968, p. 791) words, "The easiest items (conservation of number, $M = .66$) are answered correctly about twice as often as the most difficult items (distance, $M = .33$)." The data in Figure

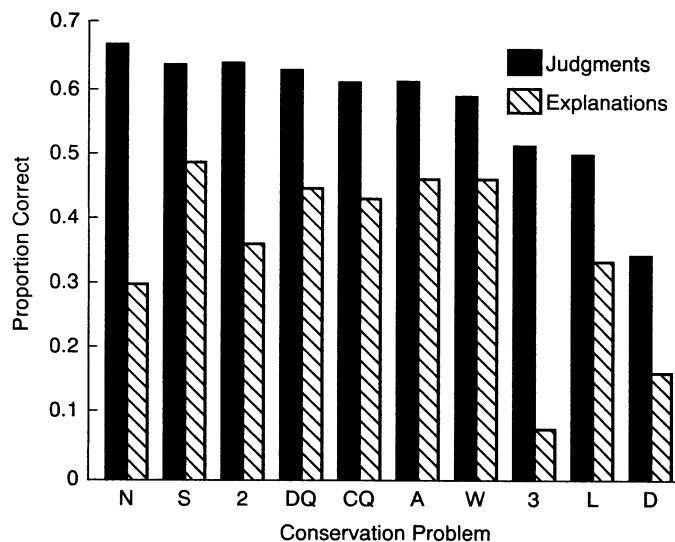


Figure 2

Proportions of correct judgments and correct explanations on Goldschmid and Bentler's (1968) 10 conservation problems. N = number, S = substance, 2 = 2-dimensional space, DQ = discontinuous quantity, CQ = continuous quantity, A = area, W = weight, 3 = 3-dimensional space, L = length, and D = distance.

2 reveal three difficulty clusters: a hard cluster composed of 1 problem (distance, 33% pass rate), an intermediate cluster composed of two problems (length and three-dimensional space, 49% and 50% pass rates), and an easy cluster composed of the remaining seven problems (58%–66% pass rates).

A third proposal that Piaget made, which at the time seemed just as improbable as the first two, is that children fail conservation problems because they do not grasp some seemingly self-evident logical principles. Five principles that were regularly mentioned in this connection, each of which is defined in Table 1, are addition/subtraction, compensation, inversion, qualitative identity, and quantitative identity. Piaget's evidence for them came from the explanations that children supplied for their judgments. On the one hand, he reported that children who gave "still same" judgments invariably justified their responses via principles of this sort. On the other hand, he reported that children who gave "different" judgments did not cite such principles and did not even seem to understand them when closely interrogated. Instead, they cited invalid rules whose common theme was that visual appearance determines quantitative relationships, such as "Longer rows always have more than shorter rows" (nonconservers of number), "Tall glasses always have more to drink than short glasses" (nonconservers of quantity), and "Things that look shorter are shorter" (nonconservers of length). As Piaget claimed, chil-

dren's understanding of the principles in Table 1 has been found to be strongly correlated with their judgments, and improving such understanding through instruction has been found to improve their judgments (Brainerd & Allen, 1971). Further, like the conservation concepts themselves, knowledge of these principles seems to follow a developmental sequence, with qualitative identity usually being understood first, followed by addition/subtraction and inversion at about the same age, then quantitative identity, and lastly by compensation (Brainerd, 1979).

NONSUPPORTIVE EVIDENCE

Although the foregoing proposals received good empirical support, they were neither the only nor the most important ones that Piaget made about conservation. Others were more fundamental because they were more intimately connected to his stage theory of cognitive development. Support for these latter proposals has been much harder to come by; indeed, some of them have been widely disconfirmed. Three will be cited here.

First, the most startling disconfirmations were concerned with Piaget's views about teaching conservation concepts to children who do not yet understand them. Early on, Piaget had been asked about whether these concepts could be taught through focused instruction. In a much-quoted remark, he dismissed this

TABLE 1
Logical principles that nonconserving children do not understand

<i>Principle</i>	<i>Definition</i>
Addition/subtraction	If $S = V$, then $S = V'$ as long as nothing was added to V or subtracted from it during $V \rightarrow V'$.
Compensation	$S = V'$ because $V \rightarrow V'$ produces compensating changes in different perceptual dimensions. For example, the transformed liquid in quantity conservation is taller, but it is narrower too.
Inversion	Following $V \rightarrow V'$, the original visual identity of the two objects could be reinstated by merely performing the reverse transformation ($V \leftarrow V'$).
Qualitative identity	Following the $V \rightarrow V'$, the transformed object is still the "same thing" (a row of chips, a glass of milk) as before.
Quantitative identity	Following $V \rightarrow V'$, the transformed object still has the "same amount" (number, length, etc.) as before.

possibility as "the American question." Later, he clarified his position as consisting of the maxims that learning is always "subject to the general constraints of the current developmental stage," that it "is no more than a sector of cognitive development that is facilitated by experience," and therefore that children's susceptibility to conservation teaching will "vary very significantly as a function of the initial cognitive levels of the children" (Piaget, 1970a, pp. 713-715). Furthermore, "teaching children concepts that they have not attained in their spontaneous development . . . is completely useless" (Piaget, 1970b, p. 30). The overriding notion that children's ability to benefit from conservation teaching is somehow constrained by their developmental stage came to be called "the stage-learning hypothesis" (Brainerd, 1977).

Researchers derived three basic predictions from the stage-learning hypothesis (Brainerd, 1978): (1) Children who show no evidence of conservation ought to be more difficult to teach than children who show at least partial understanding, because Piagetian theory says that the latter children have reached higher levels of cognitive development; (2) it should be next to impossible to teach children whose ages place them well below that for the onset of the stage at which conservation normally develops (concrete operations in Piaget's theory); (3) teaching methods that emulate the unstructured, self-discovery learning that fosters "spontaneous development" in everyday life should work better than tutorial methods that provide highly focused, laboratory-style instruction on specific concepts.

None of these predictions was borne out in experimentation. To test the first prediction, experiments were conducted in which children were first divided into groups who knew less versus more about conservation. (A typical low-knowledge group might fail all of the items on the Goldschmid-Bentler [1968] test, whereas a typical high-knowledge group might pass half the items.) Both groups then received identical instruction. With a variety of teaching methods, it was found that instruction produced equal amounts of conservation learning in low- and high-knowledge groups (Brainerd, 1977). To test the second prediction, learning experiments were conducted with preschoolers whose ages were 3 or 4 years below the nominal age range for the onset of the concrete-operational stage. Although, of course, preschoolers did not learn as rapidly as elementary schoolers, they, too, were able to learn conservation via instruction (Brainerd, 1977). To test the third prediction, "natural," self-discovery teaching methods and "artificial" tutorial methods were studied in several experiments. The most commonly used tutorial methods, which are described more fully in Table 2, were attentional training, corrective feedback, observational learning, and rule instruction. These methods produced much better conservation learning than self-discovery methods (Brainerd, 1978). On the whole, then, conservation learning experiments provided all-round disconfirmations of Piaget's stage-learning hypothesis.

A second source of nonsupportive evidence was the discovery that most claims about conservation, includ-

TABLE 2
Tutorial teaching methods for conservation concepts

<i>Teaching Method</i>	<i>Definition</i>
Attentional training	Children are taught to pay attention to the quantitative information on conservation problems and to ignore spatial information.
Corrective feedback	Incorrect judgments on conservation problems are corrected ("wrong") by the experimenter and correct judgments are rewarded ("right").
Observational learning	Nonconserving children observe a model (live or filmed) and give correct judgments and correct explanations on conservation problems.
Rule instruction	Nonconserving children are taught one or more of the logical principles in Table 1.

ing those for which there seemed to be support, were dependent on the methodological ins and outs of tests. The linguistic demands of conservation tests were the most ubiquitous sources of such task variables. Beginning with a dissertation by M. D. S. Braine (1959), many investigators conjectured that age norms and developmental sequences might look rather different if the levels of linguistic sophistication required by Piaget's tests were reduced. As noted earlier, conservation tests have two components that vary in their linguistic demands: same-different judgments (lower demands) and extemporaneous explanations (higher demands). Age norms for the acquisition of individual concepts and their sequence of development were found to fluctuate markedly as a function of whether correct judgments or correct explanations were used as the criterion for passing a conservation test (Brainerd, 1973). This can be seen in Figure 2, where proportions of correct judgments and correct explanations are displayed for Goldschmid and Bentler's (1968) ten problems. When judgment data are used (white bars), children pass more than half the problems ("still same" rate = 56%), with number conservation being the first concept to be understood and distance conservation being the last. When the explanation data are used, children pass only a third of the problems (correct explanation rate = 34%), with substance conservation being the first concept understood and three-dimensional space conservation being the last. The conservation-related logical principles in Table 2 exhibit the same dramatic variability in age norms and developmental sequences as a function of the use of judgment versus explanation criteria (Brainerd, 1979).

The third source of nonsupportive evidence comes from studies of nonconservation in adults. Piaget thought that once children had completed the concrete-operational stage of development, they regarded conservation concepts as logically certain truths. Since the age range for this stage was put at 7 to 11, this means that from early adolescence onward, people should hold these concepts with all the tenacity conferred by the force of logical certainty. In fact, they do not. G. A. Winer, R. K. Craig, and E. Weinbaum (1992) and G. A. Winer and C. McGlone (1993) have investigated this claim in many experiments using modified problems that suggest the possibility of nonconservation. For example, Winer and McGlone used

suggestive weight problems that asked, "When do you weigh more, when you are running or walking?" and suggestive number problems that asked, "Who has more, you or I?" In all of his studies, Winer has found that even adults are far from unswerving in their adherence to conservation. Instead of resisting with the unshakable confidence of logical necessity, they often succumb to suggestions of nonconservation. In Winer and McGlone's experiments, for instance, the suggestive weight and number problems were administered to college students. These subjects succumbed on 46 percent of the weight problems and on 20 percent of the number problems.

It has often been remarked that, paradoxically, the lasting contributions of great psychological theorists lie not in their theories but in the behavioral phenomena that they discover. This is certainly true in Piaget's case. While the theory itself no longer dominates the landscape as it once did, the new forms of intelligence that Piaget introduced, especially conservation concepts, continue to be regarded as critical phenomena that any credible theory of cognitive development must explain.

(See also: CHILD AND ADOLESCENT INTELLIGENCE.)

BIBLIOGRAPHY

- BRINE, M. D. S. (1959). The ontogeny of certain logical operations: Piaget's formulation examined by nonverbal methods. *Psychological Monographs* 73(5, Whole No. 475).
- BRAINERD, C. J. (1973). Judgments and explanations as criteria for the presence of cognitive structures. *Psychological Bulletin*, 79, 172-179.
- BRAINERD, C. J. (1977). Cognitive development and concept learning: An interpretative review. *Psychological Bulletin*, 84, 919-939.
- BRAINERD, C. J. (1978). Learning research and Piagetian theory. In L. S. Siegel & C. J. Brainerd (Eds.), *Alternatives to Piaget: Critical essays on the theory*. New York: Academic Press.
- BRAINERD, C. J. (1979). Markovian interpretations of conservation learning. *Psychological Review*, 86, 181-213.
- BRAINERD, C. J., & ALLEN, T. W. (1971). Experimental inductions of the conservation of "first-order" quantitative invariants. *Psychological Bulletin*, 75, 128-144.

- GOLDSCHMID, M. L., & BENTLER, P. M. (1968). The dimensions and measurement of conservation. *Child Development*, 39, 787-802.
- PIAGET, J. (1970a). Piaget's theory. In P. H. Mussen (Ed.), *Carmichael's manual of child psychology*. New York: Wiley.
- PIAGET, J. (1970b). A conversation with Jean Piaget. *Psychology Today*, 3(12), 25-32.
- PIAGET, J., & INHELDER, B. (1969). *The psychology of the child*. New York: Basic Books.
- WINER, G. A., CRAIG, R. K., & WEINBAUM, E. (1992). Adults' failure on misleading weight conservation tests: A developmental analysis. *Developmental Psychology*, 28, 109-120.
- WINER, G. A., & MCGLONE, C. (1993). On the uncertainty of conservation: Responses to misleading conservation questions. *Developmental Psychology*, 29, 760-769.

CHARLES J. BRAINERD

CHROMOSOMAL ABNORMALITIES Prior to the middle of the twentieth century, it was believed that human cells contained a total of forty-eight chromosomes and, before 1921, that sex was determined by the number of X chromosomes present in human cells. Then, two major breakthroughs in the 1950s had a significant impact on the field of medical genetics. The first came in 1956, when it was discovered that there were forty-six chromosomes in human cells, and the second in 1959, when three previously known disorders, DOWN SYNDROME, Turner syndrome, and Klinefelter syndrome, were found to be the result of an abnormal number of chromosomes. Since then, it has become apparent that chromosomal abnormalities, which are either numerical or structural in nature, play an important role in genetic disease. They are responsible for a large percentage of pregnancy loss, occasionally are found in the presence of congenital malformations, are responsible for a portion of mental retardation, and play an important role in the development of human cancer. Chromosomal abnormalities are found in about 50 percent of first-trimester spontaneous abortions. In pregnant women over 35 years of age, chromosome abnormalities are found in 2 percent of fetuses. They also are present in about 1 in 160, or 0.7 percent, live births. Chromosome analysis can be performed on any tissue, but most commonly, the study is performed on blood after birth and on

amniotic fluid cells obtained from amniocentesis during a pregnancy.

Most numerical chromosomal abnormalities result from nondisjunction, which is the failure of a pair of chromosomes to separate normally into two reproductive cells that each have twenty-three chromosomes. In numerical chromosomal abnormalities, this phenomenon is more commonly maternal in origin, but this is not the case in sex chromosomal abnormalities, where it frequently can occur in reproductive cells of either parent.

In contrast to numerical aberrations, structural chromosomal abnormalities result from breakage of one or two chromosomes, which then can be followed by reconstitution in an abnormal fashion (Figure 1). This rearrangement can lead to duplication or deletion of a portion of a chromosome(s), with significant physical and mental consequences. Most commonly, the rearrangement of chromosomal material between two chromosomes occurs in a balanced fashion with no loss or gain of genetic material. The affected individual, therefore, does not have any adverse effects from this rearrangement; but there is a significant risk for unbalanced chromosome material (i.e., extra or missing chromosome segments) in the affected individual's offspring, which usually leads to pregnancy loss. Balanced structural abnormalities occur in approximately 1 in 500 newborns, and balanced rearrangements occur about three times more frequently than do unbalanced ones.

Advances in cytogenetic technology since the 1960s have greatly expanded our knowledge of the role that chromosomal abnormalities play in mental retardation. Although the mechanism leading to abnormal development of the central nervous system is not known, mental retardation is a constant feature in numerical and unbalanced structural chromosomal abnormalities not involving the sex chromosomes.

The ability to stretch chromosomes through new staining techniques and to combine these studies with new molecular genetic techniques has resulted in the identification of subtle chromosomal abnormalities in a number of known conditions previously considered to be of undetermined origin. This is exemplified by the discovery of a subtle deletion of chromosome material on one number 15 chromosome in 50-70 percent of individuals with Prader-Willi syndrome, a dis-

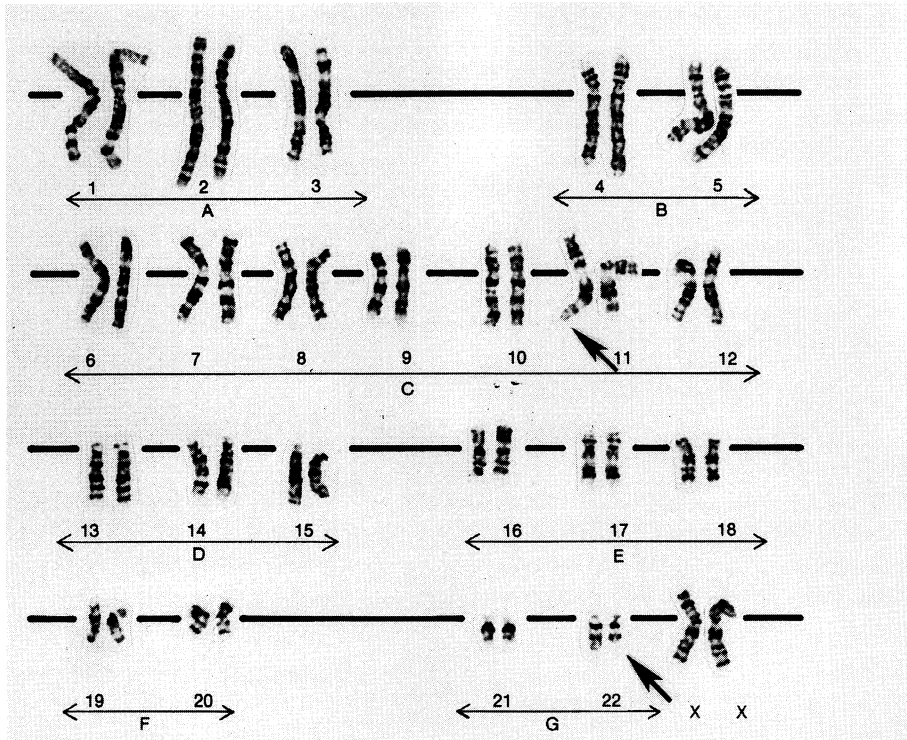


Figure 1

Balanced translocation between the long arm of one chromosome 11 and the long arm of one chromosome 22

order that results in obesity, hypotonia (low muscle tone), mental retardation, hyperphagia (intensification of appetite), hypogonadism, and short stature. Using DNA technology, researchers have determined that the deletion is always paternally derived. In Angelman syndrome, by contrast, another recognizable disorder that results in mental retardation, neurological abnormalities, and seizures, a deletion in the same region of the number 15 chromosome is present in 80 percent of cases, and it is always maternally derived. Even in cases in which there is no discernible chromosomal abnormality, the two normal number 15 chromosomes in Prader-Willi syndrome are both of maternal origin, while the two number 15 chromosomes in Angelman syndrome occasionally are of paternal derivation. These findings emphasize the importance of receiving an equal contribution of chromosomes from both par-

ents, or genomic imprinting, to enable the fetus to develop normally, both physically and mentally.

With the above information representing a brief introduction to chromosomes and the types of chromosome abnormalities, we will now discuss a few of the more common chromosome syndromes and their associated physical and cognitive characteristics.

AUTOSOMAL CHROMOSOMAL ABNORMALITIES

Down Syndrome. Down syndrome results from an extra number 21 chromosome (trisomy 21). The condition is the most common genetic cause of mental retardation. It occurs in approximately 1 in 800 live births, and its incidence increases with advanced maternal age (above 35 years). Of those conceptions

that are the result of trisomy 21, about three-fourths do not survive to term. In this condition, first described by Langdon Down in 1866, affected individuals have a characteristic appearance. Facially, the eyes are upslanting, with flattening of the bridge of the nose and extra skin folds located over the medial aspect of the eyes. The ears are small, and typically the top of the ear is overfolded. The tongue is large and frequently protrudes out of the mouth. Increased skin over the posterior aspect of the neck is common, and occasionally, during pregnancy, can be identified in an affected fetus by ultrasound. The hands tend to be broad, with short fingers, and the fifth fingers frequently are curved inward. Abnormalities of the creases of the palm are common, and frequently only one crease extends across the entire palm, rather than two shorter creases typically seen in an unaffected individual. Increased distance between the first and second toes can be observed, with a deep skin groove extending along a portion of the sole of the foot. Approximately 40 percent of affected infants have congenital heart disease. Intestinal obstruction occurs in 12 percent, and congenital cataracts in another 3 percent of affected neonates. Hypothyroidism occurs more frequently. Although serious health problems do not develop in most children with Down syndrome, leukemia develops in approximately 1 percent of cases, which is about fifteen times more frequent than in children without Down syndrome. Close to 15 percent of affected children have abnormal alignment of the cervical spine, which could lead to spinal cord compression and neurological damage following a significant traumatic injury to the neck.

Hypotonia, which is associated with joint laxity, is a constant feature at birth, and developmental delay is apparent in the first year of life. At that time, infants with Down syndrome usually exhibit significant delays in their cognitive and motor development, and by school age, most demonstrate moderate to severe degrees of mental retardation, with IQs between 30 and 50 on most standardized tests of intelligence. It has been shown that children with Down syndrome have relative delays in the development of language skills, even when compared with children with comparable mental ages. Their communication difficulties may be complicated further by significant speech articulation problems. An area of strength often seen in affected

individuals is their social and adaptive skills, providing them with the potential to procure some form of meaningful employment in adulthood and adequate functioning in a semi-independent living setting.

Survival to adulthood is now common, with close to 50 percent surviving to 50 years of age and about 1 in 7 to 68 years of age. Although affected males typically are sterile, fertility can be present in females with Down syndrome. The longer survival rate among individuals with Down syndrome has drawn greater attention to premature aging and the presence of the typical neuroanatomic changes of Alzheimer's disease by 35 years of age, which can lead to premature senility.

Ninety-five percent of Down syndrome results from trisomy 21, which is usually maternally derived, although both parents have normal chromosomes. In another 4 percent, a translocation occurs, with the extra number 21 chromosome usually located on the top of one of the number 14 chromosomes. In approximately 40 percent of translocations, a parent is a silent carrier for the abnormality, necessitating chromosome studies on the parents. If the mother is found to be a carrier, recurrence risk in future offspring is in excess of 10 percent; if it is paternal in origin, recurrence risk is less than 5 percent. These risks are in contrast to trisomy 21, in which risk to future offspring is approximately 1 percent. Mosaic Down syndrome, which occurs in about 1 percent of cases, is the presence in the affected individual of two populations of cells, one with three number 21 chromosomes and the other with two number 21 chromosomes. The presence of mosaicism, which develops in the first few cell divisions following a conception with trisomy 21, can modify the consequences of the extra number 21 chromosome, potentially resulting in higher intellectual functioning.

Cri-du-Chat Syndrome. In the neonatal period, the presence of a meowl-like cry in a young infant usually enables recognition of Cri-du-Chat syndrome, which is a rare condition but a common structural chromosomal abnormality. The chromosome abnormality occurs in approximately 1/20,000 live births. Clinical manifestations result from a missing portion of the short arm of one of the number 5 chromosomes (Figure 2). In addition to the unusual cry, affected infants have a small head size, round face,

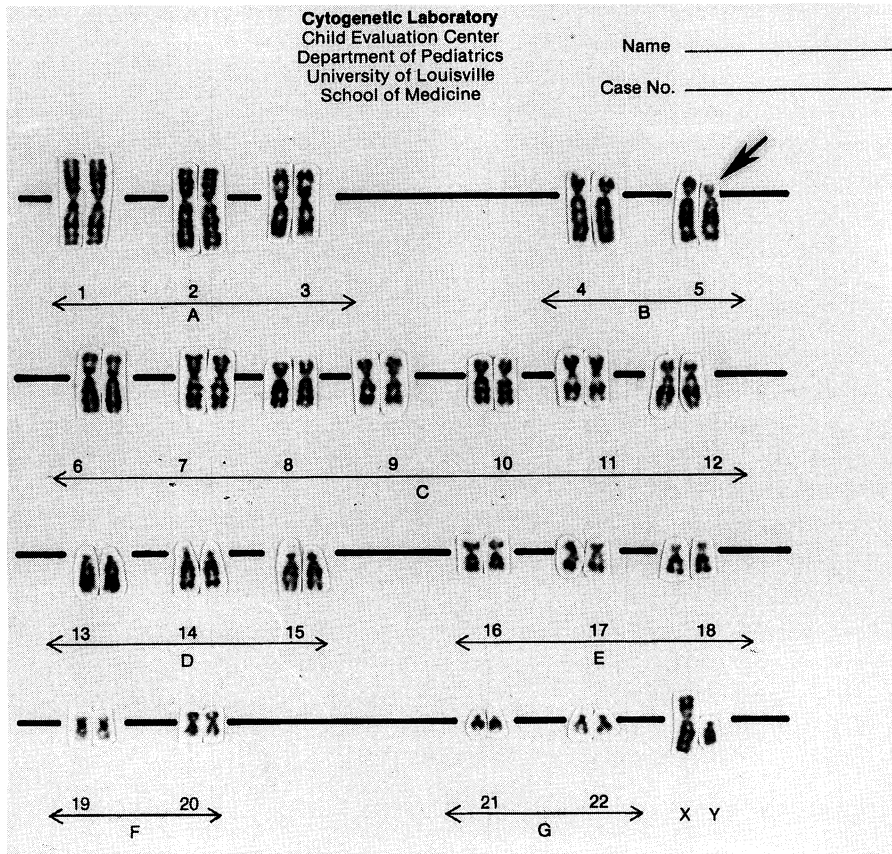


Figure 2

Terminal deletion of chromosomal material on the short arm of one chromosome 5

downslant to the eyes, and a small chin. Feeding difficulties are common in infancy, and slow growth and muscle-tone abnormalities are frequently present. Intellectual abilities are usually in the severe range of mental retardation, although some variability may be observed. Language skills are severely delayed, even relative to an affected individual's intellectual abilities. Hyperactivity and a short attention span are common, and frequently the child appears to be in constant motion.

In about 85–90 percent of cases, the number 5 deletion occurs spontaneously, either in an egg or sperm at the time of conception. Nonetheless, chromosome studies are necessary on the parents of an affected infant, because the presence of a balanced chromosomal rearrangement between one of the number 5 chromosomes and another chromosome in one parent is found in 10–15 percent of cases, substantially increasing the risk for recurrence.

SEX CHROMOSOME ABNORMALITIES

Sex chromosome abnormalities also can be numerical or structural in nature, and the abnormality can be present either in all cells or in only a portion, resulting in mosaicism. X and Y chromosomal aneuploidy (extra or missing sex chromosomes) is relatively common, with an overall frequency of 1 in 500 births. In general, clinical manifestations, including developmental problems, are less severe in comparison to autosomal chromosomal abnormalities (i.e., alterations involving chromosomes 1–22). An increased risk for sterility exists regardless of whether the abnormality occurs in a male or a female. Generally, there is less of an age-related risk for sex chromosomal abnormalities, unlike Down syndrome.

Sexual differentiation is determined by the presence or absence of a Y chromosome. Genes on the Y chromosome direct normal testicular differentiation. In a

male with an extra X chromosome, which is seen in Klinefelter syndrome, normal testicular development is altered. When a Y chromosome is not present, development of the female reproductive tract occurs. If the fetus is a female and either one X chromosome is absent or a portion of it is altered, the result is Turner syndrome.

Klinefelter Syndrome. In Klinefelter syndrome, a sex chromosome abnormality that is found in 1 in 1,000 live-birth males, there is an extra X chromosome (47, XXY). In 85 percent of cases, the extra X chromosome is found in all cells, but mosaicism exists in the remainder. Affected males have a normal physical appearance, although they are somewhat taller than chromosomally normal males. XXY males generally appear to be physically normal until puberty, when secondary sexual development is incomplete. The testes remain small and poorly developed, and in addition to underdevelopment of secondary sexual characteristics, infertility is a constant feature. Intelligence usually is only slightly impaired, although in at least two-thirds, greater verbal deficits are present, leading to problems particularly in language-based academic areas. Poor psychosocial development also is frequently observed. Use of testosterone in early adolescence not only has a beneficial effect on pubertal changes but may also enhance psychosocial skills.

Turner Syndrome. Unlike males with sex chromosome aneuploidy, females with Turner syndrome are usually recognizable at birth by the presence of physical abnormalities. Although Turner syndrome accounts for approximately 1.5 percent of all conceptions, up to 99 percent are lost spontaneously in utero, accounting for about 18 percent of all chromosomally abnormal spontaneous abortions. The missing X chromosome is more commonly paternal than maternal in origin. Many females with Turner syndrome are small at birth and continue to grow at a significantly slower rate in mid and late childhood, resulting in an adult stature of 54–56 inches; use of growth hormone can lead to taller stature. Structural abnormalities of the cardiovascular system and kidneys and webbing of the neck are occasionally present. In utero, the ovaries have a normal appearance, but there is rapid programmed destruction in affected females, leading to fibrous streaks after birth, resulting in infertility and

absence of secondary sexual characteristics. In early adolescence, the use of sex hormones will achieve pubertal changes and menses, enhancing an affected female's body image.

In about 60 percent of cases, one X chromosome is missing. In the other 40 percent, other abnormalities of the X chromosome are found that may mitigate the phenotype but not the lack of reproductive potential. For the majority of individuals with Turner syndrome, the overall intellectual level falls within the average range. Strengths are usually in the verbal modality, and weaknesses, in spatial direction, perceptual-motor ability, and fine-motor skills. It is very common for affected individuals to have particular difficulties with arithmetic reasoning and mathematical computation.

Fragile X Syndrome. Significant mental retardation occurs more commonly in males than in females. In 1969, researchers discovered the fragile X abnormality by using a special growth-promoting solution deficient in folic acid in the preparation of karyotypes (chromosome analysis) on family members with a hereditary form of mental retardation in which there was a predominance of affected males. The abnormality is characterized by a constricted area at the bottom of the X chromosome, which occasionally breaks off of the chromosome itself (Figure 3). Fragile X syndrome is second to Down syndrome among all chromosomal causes of significant mental retardation in males and is the most common heritable form of mental retardation. In males, fragile X syndrome occurs in about 1 in 1,500 live births. Mental retardation, which usually is of a significant degree, occurs in 80 percent of affected males. Although physical features may be normal, facial abnormalities are frequently observed, notably an elongated face, prominent jaw, and large ears. In adult males, testicular size is enlarged. Specific speech and language problems are present, and features of an autistic disorder occasionally are observed. In all of these individuals, the fragile X chromosome can be identified in up to 40 percent of cells counted on a chromosome analysis performed on a blood specimen. The incidence of fragile X syndrome in females is about 1 in 2,500. In two-thirds of females with the fragile X abnormality, no intellectual deficits are present, but in the other one-third, deficits in

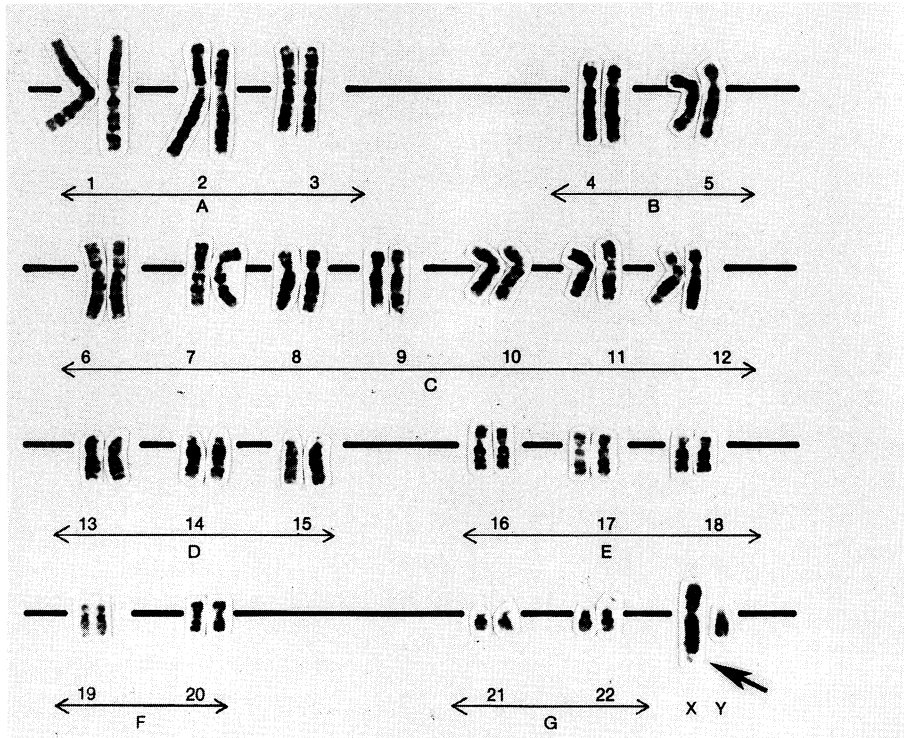


Figure 3

Fragile X site on the long arm of one X chromosome

cognitive abilities, including mental retardation, are present. Although identification of the fragile X chromosome is common in a blood specimen of a female with fragile X syndrome who has intellectual impairments, it is often not apparent in the fragile X female who has normal cognitive abilities.

Unlike other genetic disorders involving the X chromosome, in which affected males always have clinical manifestations, no clinical abnormalities or intellectual deficits occur in 20 percent of males (non-expressing, transmitting males); but the individuals transmit the genetic abnormality to all of their female offspring. Although their daughters are intellectually normal, these women are at risk for having affected male offspring with mental retardation. The reason why some fragile X males are intellectually normal was not understood until 1991, when the gene causing

fragile X syndrome was discovered. It was found that a region of the fragile X gene normally varies in size. As this portion of the gene expands, the risk for mental retardation increases. Therefore, the larger the size of this region of the gene, the greater the risk for altered gene expression leading to mental retardation.

In the future, through the combined use of cytogenetic and molecular technology, our understanding of the importance that cytogenetic abnormalities play in altering human function should continue to grow. It is anticipated that the knowledge that is gained from these discoveries will provide insights into new treatment modalities for conditions that significantly alter normal physical and mental development.

(See also: GENETICS, BEHAVIOR.)

BIBLIOGRAPHY

- GARDNER, R. J. M., & SUTHERLAND, G. R. (1989). *Chromosome abnormalities and genetic counseling*. New York: Oxford University Press.
- GELEHRTER, T. D., & COLLINS, F. S. (1990). *Principles of medical genetics* (pp. 159–189). Baltimore, MD: Williams & Wilkins.
- HOOKE, E. B., & HAMERTON, J. L. (1977). The frequency of chromosome abnormalities detected in consecutive newborn studies, differences between studies, results by sex and severity of phenotypic involvement. In *Population cytogenetics: Studies in humans* (pp. 63–79). New York: Academic Press.
- LYON, M. F. (1988). X chromosome inactivation and the location and expression of X-linked genes. *American Journal of Human Genetics*, 42, 8–16.
- RATCLIFFE, S. G., & PAUL, N. (EDS.). (1986). *Prospective studies on children with sex chromosome aneuploidy*. March of Dimes Birth Defects Foundation, Birth Defects Original Article Series 22(3). New York: Alan R. Liss, Inc.
- SUTHERLAND, G. R., HAAN, E. A., & KREMER, E. (1991). Hereditary unstable DNA: A new explanation for some old genetic questions? *Lancet*, 388, 289–292.
- THOMPSON, M. W., MCINNES, R. R., & WILLAR, H. F. (1991). *Thompson and Thompson genetics in medicine* (4th ed., pp. 201–246). Philadelphia: W. B. Saunders.

JOSEPH H. HERSH
ALLAN BLOOM
MARGARET J. BARCH

CLASSIFICATION OF INTELLIGENCE

Individual differences in human ability have been acknowledged since the beginning of recorded history. For example, as early as 2200 B.C. Chinese emperors used an elaborate system of competitive examinations to select personnel for government positions (DuBois, 1970). Not until the 1800s did scientists and educators begin to evolve a vocabulary to describe the spectrum of ability. Two landmarks are pertinent in this regard. In 1866, the French physician O. Edouard Séguin published *Idiocy, and Its Treatment by the Physiological Method*, the first major textbook on what is now termed mental retardation or developmental disability. In 1869, British scientist Francis GALTON published *Hereditary Ge-*

nius, in which he argued that genetic factors were maximally important for the attainment of eminence. Albeit a crude beginning, by referring to the lower extreme of functioning as idiocy and the higher extreme of functioning as genius, these authoritative sources heralded the era of classification.

Usable systems for the classification of intelligence arose with the invention of the individual intelligence test in the early 1900s. Dozens of classification systems have been proposed since then. The surplus of classification schemes is, in part, due to the release of new tests but also reveals the evolution of concepts in the field. The purpose of this commentary is to review historical trends in classification, describe the continuum of intelligence, survey the cutoff points used with contemporary tests, and consider problems that arise in the classification of intelligence. We begin with a brief history of approaches to classification.

A BRIEF HISTORY OF
INTELLIGENCE CLASSIFICATION

Based upon statistical concepts rather than formal tests, Galton (1869) proposed sixteen classifications of ability in his investigation of hereditary genius. The seven classes *a* through *g* represented increasingly less capable (and less frequent) varieties of below-average ability; the eighth class *x* (all grades below *g*) was also used. The seven classes *A* through *G* signified increasingly more capable (and less frequent) varieties of above-average ability; the eighth class *X* (all grades above *G*) was reserved for the exceptional genius, observed once in a million cases. Galton (1869) described four of the superior classes as follows:

The class C possesses abilities a trifle higher than those commonly possessed by the foreman of an ordinary jury. D includes the mass of men who obtain the ordinary prizes of life. E is a stage higher. Then we reach F, the lowest of those yet superior classes of intellect, with which this volume is chiefly concerned (p. 31).

Galton believed that the distribution of ability was symmetrical, with “gifted” men possessing a level of intelligence raised as much above the average level as “idiots” were depressed below it. Since his classification scheme was merely descriptive and not based upon test results, it quickly faded from view.

Empirical systems for classification began to appear shortly after 1905, when Alfred Binet and Théophile Simon created the individual intelligence test. Following that extraordinary contribution to the objective measurement of intelligence, Henry Goddard (1910a) translated the Binet-Simon scales from French into English and set about testing and categorizing thousands of children in U.S. schools and residential institutions. In 1911, Goddard published a simple, unadorned classification of intelligence (Figure 1). Based upon the test results for 1,547 schoolchildren, the graph depicted the proportion whose tested mental age (MA) was at, above, or below their chronological age (CA). Goddard used four categories—those with MA 4 years or more behind CA were termed *feeble-minded* (3%); those with MA 2 and 3 years behind CA were termed *merely backward* (15%); those at or within 1 year of their CA were termed *normal* (78%), and those with MA more than 1 year above CA were termed *gifted* (4%).

In 1912, Wilhelm Stern suggested that an intelligence quotient computed from MA divided by CA would give a better measure of the relative functioning of a child as compared with same-aged peers. A few years later, Lewis TERMAN (1916) extended Stern's idea and formally proposed the use of the intelligence quotient (IQ) concept: $IQ = MA \div CA \times 100$. From

this point, classification systems were based, at least partly, on IQ or a similar global score.

A special concern with the early systems was the classification of persons functioning in the low end of the intelligence spectrum. As early as 1877, well before the advent of individual tests, three broad categories of deficient intellectual functioning were recognized clinically—*idiot*, *imbecile*, and *feeble-minded* (Barr, 1904; Scheerenberger, 1983; Wilmarth, 1906). Goddard (1910b) invented the diagnostic label *moron* (from the Greek *moronia*, meaning “foolish”) as a replacement for the feeble-minded category. Shortly thereafter, Lewis Terman (1916) of Stanford University published a new test, the Stanford-Binet, and popularized the tripartite division of idiot, imbecile, moron in the first comprehensive numerical classification of intelligence by differing levels of IQ (Table 1).

During the early and mid-1900s, new individual intelligence tests, such as the various Wechsler scales, were developed to compete with successive editions of the Stanford-Binet (Terman & Merrill, 1937, 1960; Wechsler, 1939, 1944, 1949, 1955). With each new test or revision, the developers proposed a slight variation on Terman's 1916 seminal classification scheme. Because such individuals would be found only once out of every million persons, the term *genius* gradually disappeared from most classification systems. Also,

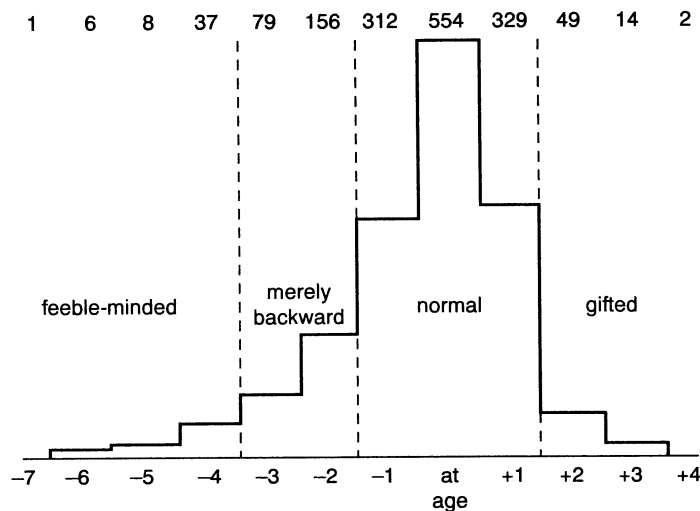


Figure 1
 The distribution of mental age versus chronological age for 1,547 U.S. schoolchildren
 SOURCE: Goddard, H. H. (1911).

TABLE 1
Classification of intelligence based upon the
Stanford-Binet (1916)

IQ	Classification
Above 140	Near genius or genius
120–140	Very superior intelligence
110–120	Superior intelligence
90–110	Normal, or average intelligence
80–90	Dullness, rarely classifiable as feeble-mindedness
70–80	Borderline deficiency, sometimes classifiable as dullness, often as feeble-mindedness
50–70	Morons
25 ^a –50	Imbeciles
Below 25 ^a	Idiots

^aAn indefinite cutoff of 20–25 was specified.

SOURCE: Terman, L. M. (1916), *The measurement of intelligence*.

cutoffs for the various categories were shifted slightly from time to time. Nonetheless, the categories proposed by Terman (1916) and shown here in Table 1 held sway until the 1950s.

Alongside developments in testing, authorities in the field of mental retardation proposed new labels to replace the terms idiot, imbecile, and moron—which became increasingly pejorative. For a few decades, chaos reigned in classification terminology. By the late 1950s, some twenty-three systems for the classification of deficient intelligence had been proposed (Gelof, 1963). These systems differed in the overall generic term for mental retardation (e.g., intellectually handicapped, mentally defective, mentally handicapped, slow learner, scholastically retarded, educationally subnormal) and also proposed a variety of cutoff points for retarded functioning (e.g., IQ below 65, or 70, or 75, or 85).

In 1961, a new system proposed by the American Association on Mental Deficiency (AAMD) gained prominence (Heber, 1961). Mental retardation was designated as *mild*, *moderate*, *severe*, or *profound*, based upon the concurrently measured degree of impairment in intellectual functioning and adaptive behavior. The AAMD specified IQ ranges for the various levels of retardation but also insisted on the need to assess adaptive behavior along with IQ. In state and federal statutes, as well as in everyday professional practice, a

deficit in adaptive behavior is now considered an essential component of mental retardation (Grossman, 1983). The AAMD is now known as the AAMR (American Association on Mental Retardation).

AN OVERVIEW OF LEVELS OF FUNCTIONING

As a starting point for describing levels of functioning, we consider a prominent classification of intelligence based on the Wechsler Adult Intelligence Scale-Revised (WAIS-R) (Table 2). The lower levels of functioning have been extended along the lines proposed by the AAMR (Grossman, 1983) and endorsed by the American Psychiatric Association (APA, 1987). What is the merit of such a classification system?

Although IQ is commonly overvalued by those who do not understand its limitations, it is nonetheless true that overall performance on the WAIS-R and similar tests of intelligence for children and adults is a good predictor of the potential for various social, educational, and occupational attainments (Table 3). This is especially true in the lower ranges of the spectrum. There is a large body of research that has shown that low IQ exerts a pervasive influence on the cognitive and adaptive (i.e., academic and socio-occupational)

TABLE 2
Classification of intelligence based upon the
WAIS-R (1981) and AAMR criteria (1983)

IQ	Classification
Above 129	Very superior
120–129	Superior
110–119	High average
90–109	Average
80–89	Low average
70–79	Borderline
55–69 ^a	Mild mental retardation
40–54	Moderate mental retardation
25–39	Severe mental retardation
Below 39	Profound mental retardation

^aThe cutoffs for the four levels of retardation are inexact: below *approximately* 70 (mild), below 50–55 (moderate), below 35–40 (severe), and below 20–25 (profound).

Furthermore, a concurrent deficit in adaptive behavior and onset before age 18 is also required for the classification of mental retardation.

TABLE 3
IQ levels and typical social, educational, and occupational attainments

IQ	Criteria
125–130	Mean of persons receiving medical, doctorate, and law degrees; mean for technical degrees such as chemistry and engineering
115–120	Mean of college graduates; mean for professional positions such as teachers, managers, and policemen
105–110	Mean of high school graduates; 50:50 chance of graduating from college
100	Average for the general population; about half of all persons score in the 90 to 109 range
90–95	Mean of persons from inner city, low income, and rural settings
85–90	Mean of unskilled laborers and persons employed for simple assembly line work
75–80	About 50:50 chance of reaching high school; need special education services; difficulty coping with modern technological society
65–70	Mild mental retardation likely; may achieve social and vocational adequacy with special training and supervision
55–60	Mild mental retardation definite; partial independence in living arrangements; may achieve fourth grade academic level
45–50	Moderate mental retardation likely; function well in a sheltered workshop; need supervised housing
35–40	Severe mental retardation likely; little or no communication skills; sensory and motor impairments; trainable in basic health habits
Below 25	Profound mental retardation likely; minimal functioning and incapable of self-maintenance; need constant nursing care and supervision

competence of individuals (Maloney & Ward, 1979; Matarazzo, 1972).

In the higher ranges of the spectrum, a particular IQ can be viewed as improving the odds that a person can manage the intellectual demands of various occupations and endeavors. Table 3 is an idealized chart that relates IQ to various social, educational, and occupational achievements. The scores depicted are based upon summaries of published studies and the

judgment of experts on intelligence (Gregory, 1992; Jensen, 1980; Matarazzo, 1972). The scores are averages only—there is considerable spread above and below these points for the individuals who comprise each of these specific groups. Furthermore, a particular IQ is only an indicator, not a sufficient condition for the attainment of a specific educational or occupational goal. The outcomes listed in Table 3 also represent probabilities, not certainties. For example, there is no reason why an individual with IQ of 91 cannot attend college and, in fact, many do. However, such a student would function at a distinct disadvantage compared to classmates and would find the educational demands to be extremely taxing. All other factors held constant, a student with IQ of 91 would be substantially less likely to graduate from college than a student with IQ of 121. However, students with an IQ of 91 who have good study habits and who put in the extra study time each day, do graduate from college each year. Additionally, for a variety of reasons, and especially motivational and behavioral, a not insignificant number of students with IQs above 120 do not graduate from college, proving once again that IQ and adaptive competence are highly interrelated but not equivalent traits.

CONTEMPORARY CLASSIFICATION SYSTEMS

An early discovery in the history of intelligence testing was that the distribution of test scores for the general population approximated a symmetrical, mathematically defined, bell-shaped curve called the normal distribution (Gregory, 1992). This was true even for tests constructed entirely without reference to the normal curve. Subsequently, by 1950 test developers used statistical sampling procedures in choosing representative groups of individuals upon whom to develop their intelligence tests. Careful sampling *guaranteed* that test scores fit the expected pattern of a near normal distribution.

One advantage that followed from the universal finding that IQ scores approximate a normal distribution is that the normal curve has useful mathematical properties that simplify classification. Specifically, in a normal distribution 50 percent of the test scores cluster around their own average within a distance of plus

or minus two-thirds of a standard deviation. Most IQ tests were normed so as to have a mean of 100 and a standard deviation of 15. Two-thirds of a standard deviation is therefore an elegant, handy 10 IQ points. Thus, for an IQ test with a near normal distribution, it follows that approximately 50 percent of the test scores will fall in the 20 point IQ range of 90 through 109. Classifying 50 percent of the general population as average is arbitrary but appealing in its simplicity. Likewise, as is shown in Table 2, the use of 10 point IQ ranges to bracket additional levels of above and below average functioning is also arbitrary, but firmly entrenched in modern testing practices.

In Table 4, the reader will find a summary of classification systems for the most recently developed individual intelligence tests. All these tests, except the Stanford-Binet: Fourth Edition (SB-IV), use a 20 point range to bracket average performance, and 10 point ranges to classify successive levels of functioning above

and below the average. The SB-IV uses different IQ ranges for classification because, unlike the more common use of 15 as the standard deviation, this test has a slightly larger standard deviation of 16 IQ points. Notice, too, that labels differ, especially for the lower IQ classifications. At the lowest level, IQ below 70 (68 on the SB-IV), *mentally retarded* is the term used on the WAIS-R and SB-IV, whereas the K-ABC uses the phrase *lower extreme*. The WISC-III employs the term *intellectually deficient* to avoid the implication that a very low IQ score is sufficient evidence for the classification of mental retardation.

PROBLEMS IN THE CLASSIFICATION OF INTELLIGENCE

Although classification is widely used in individual intelligence testing, a number of problems and cautions deserve brief mention. The first has to do with

TABLE 4
Ability classifications, IQ ranges, and percent of norm sample for contemporary tests

WAIS-R (1981) SD = 15		K-ABC (1983) SD = 15		SB-IV (1986) SD = 16		WISC-III (1991) SD = 15		Theoretical
Very Superior 130+	2.6%	Upper Extreme 130+	2.3%	Very Superior 132+	2.5% ^a	Very Superior 130+	2.1%	
Superior 120-129	6.9%	Well Above Average 120-129	7.4%	Superior 121-131	7.0% ^a	Superior 120-129	8.3%	6.7%
High Average 110-119	16.6%	Above Average 110-119	16.7%	High Average 111-120	16.0% ^a	High Average 110-119	16.1%	16.1%
Average 90-109	49.1%	Average 90-109	49.5%	Average 89-110	50.0% ^a	Average 90-109	50.3%	50.0%
Low Average 80-89	16.1%	Below Average 80-89	16.1%	Low Average 79-88	16.0% ^a	Low Average 80-89	14.8%	16.1%
Borderline 70-79	6.4%	Well Below Average 70-79	6.1%	Slow Learner 68-78	6.0% ^a	Borderline 70-79	6.5%	6.7%
Mentally Retarded below 70	2.3%	Lower Extreme below 70	2.1%	Mentally Retarded below 68	2.5% ^a	Intellectually Deficient below 70	1.9%	2.2%

NOTE: Theoretical percentages in far right column are based upon the normal distribution and apply to all tests except the SB-IV.

^a Percentages for the SB-IV are extrapolated from a percentile chart and therefore approximate.

WAIS-R: Wechsler Adult Intelligence Scale-Revised

SB-IV: Stanford-Binet: Fourth Edition

K-ABC: Kaufman Assessment Battery for Children

WISC-III: Wechsler Intelligence Scale for Children-Third Edition

the potential permanence of pejorative labels. There is a tendency for labels such as *borderline*, *dull normal*, and *intellectually deficient* to permanently stigmatize, because they tend to stick to individuals regardless of changes in functioning or circumstances. Psychologists, educators, and parents must remember that classification is intended to be merely a shorthand form of communication about current functioning. The terms used should not become permanent labels attached to examinees for the indefinite future. Partly in response to a concern about the stickiness of diagnostic terms, the most recent AAMR classification system downplays the use of labels (American Association on Mental Retardation, 1992).

A second concern is whether classification systems are merely arbitrary as opposed to "cutting nature at its joints." Consider the definition of mental retardation, which refers to jointly assessed deficits in adaptive behavior and low IQ. The usual IQ cutoff for mental retardation is a score below approximately 70. But why is an IQ of 70 considered the cutoff, rather than a score of 50, 60, or 80? Historical perspective reveals that the cutoff score of 70 is not arbitrary, but represents a meaningful social, education, and occupational *threshold* (Gregory, 1987). Jensen (1980) puts it this way:

To be sure, these thresholds are statistical and represent only differing probabilities for individuals' falling on either side of the threshold. But the differential probabilities are not negligible. Such probabilistic thresholds of this type occur in different regions of the IQ scale, not by arbitrary convention or definition, but because of the structure of the educational and occupational systems of modern industrial societies and their correlated demands on the kind of cognitive ability measured by IQ tests (p. 114).

An IQ of approximately 70—which, as is shown in Tables 1 through 4, is the typical cutoff point for mental retardation—represents what the experience of teachers reveals is the threshold for mastering the traditional subject matter of elementary school. It also represents the threshold for personal independence and nonsheltered employment, in the probabilistic sense described above. Thus, classification systems are not entirely arbitrary, especially in the lower regions of functioning. Nevertheless, at higher levels of func-

tioning, classification becomes relatively less important in terms of ordinary educational and occupational aspirations.

Finally, the distribution of IQ is not always as smooth, symmetrical, and near normal as represented by common classification systems. This is especially true in the very lowest ranges, up to an IQ of 50 or 55. When institutionalized persons are taken into account, most normative studies reveal a bulge of cases in the severely and profoundly retarded ranges that challenges the normal curve (Gregory, 1987).

BIBLIOGRAPHY

- AMERICAN ASSOCIATION ON MENTAL RETARDATION. (1992). *Mental retardation: Definition, classification, and systems of supports* (9th ed.) Washington, DC: AAMR.
- AMERICAN PSYCHIATRIC ASSOCIATION. (1987). *Diagnostic and statistical manual of mental disorders* (3rd ed.—rev.). Washington, DC: APA.
- BARR, M. (1904). Classification of mental defectives. *Journal of Psycho-Asthenics*, 9, 29–38.
- BINET, A., & SIMON, T. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux. *Année Psychologique*, 11, 191–244.
- DUBOIS, P. H. (1970). *A history of psychological testing*. Boston: Allyn & Bacon.
- GALTON, F. (1869). *Hereditary genius: An inquiry into its laws and consequences*. London: Macmillan.
- GELOF, M. (1963). Comparison of systems of classification relating degree of retardation to measured intelligence. *American Journal of Mental Deficiency*, 68, 297–317.
- GODDARD, H. H. (1910a). A measuring scale for intelligence. *The Training School*, 6, 146–155.
- GODDARD, H. H. (1910b). Four hundred feebleminded children classified by the Binet method. *Pedagogical Seminary*, 17, 387–397.
- GODDARD, H. H. (1911). Two thousand normal children measured by the Binet measuring scale of intelligence. *Pedagogical Seminary*, 28, 232–259.
- GREGORY, R. J. (1987). *Adult intellectual assessment*. Boston: Allyn & Bacon.
- GREGORY, R. J. (1992). *Psychological testing: History, principles, and applications*. Boston: Allyn & Bacon.
- GROSSMAN, H. J. (Ed.). (1983). *Classification in mental retardation*. Washington, DC: American Association on Mental Deficiency.

- HEBER, R. A. (1961). A manual on terminology and classification in mental retardation (rev. ed.). *Monograph Supplement American Journal of Mental Deficiency*, 64.
- JENSEN, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- MALONEY, M. P., & WARD, M. P. (1979). *Mental retardation and modern society*. New York: Oxford University Press.
- MATARAZZO, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). Baltimore: Williams & Wilkins.
- SCHEERENBERGER, R. C. (1983). *A history of mental retardation*. Baltimore: Brookes.
- SÉGUIN, O. E. (1866). *Idiocy and its treatment by the physiological method*. New York: William Wood.
- STERN, W. L. (1912). Über die psychologischen Methoden der Intelligenzprüfung. (English translation by G. M. Whipple [1914]. The psychological methods of testing intelligence. *Educational Psychology Monographs*, 13, Baltimore: Warwick & York.)
- TERMAN, L. M. (1916). *The measurement of intelligence*. Boston: Houghton Mifflin.
- TERMAN, L. M., & MERRILL, M. A. (1937). *Measuring intelligence*. Boston: Houghton Mifflin.
- TERMAN, L. M., & MERRILL, M. A. (1960). *Stanford-Binet Intelligence Scale: Manual for the third revision, Form L-M*. Boston: Houghton Mifflin.
- WECHSLER, D. (1939). *The measurement of adult intelligence*. Baltimore: Williams & Wilkins.
- WECHSLER, D. (1944). *The measurement of adult intelligence* (3rd ed.). Baltimore: Williams & Wilkins.
- WECHSLER, D. (1949). *Manual for the Wechsler Intelligence Scale for Children*. New York: Psychological Corporation.
- WECHSLER, D. (1955). *Manual for the Wechsler Adult Intelligence Scale*. New York: Psychological Corporation.
- WILMARTH, A. (1906). To whom may the term, feeble-minded, be applied? *Journal of Psycho-Asthenics*, 10, 203–205.

ROBERT J. GREGORY

COGNITION See INFORMATION PROCESSING.

COGNITIVE STYLES From its post-World War II origins until the present day, the domain of cognitive style has had to face the question of the extent and nature of its link to intelligence. This concern

is understandable because several cognitive styles have abilitylike properties; hence it has been important to know whether the newer cognitive-style constructs offered something unique to the field. Before addressing this concern directly, however, it may prove helpful to consider how and why cognitive styles emerged on the psychological scene in the first place.

Cognitive styles reflect individual differences in perception, attention, memory, decision making, judgment, and concept attainment. These represent some of the standard categories of experimental psychology, but consistent with L. J. CRONBACH'S (1957) observations on the two psychologies of the time, experimental psychologists had little, if any, interest in individual differences in cognitive functioning. A small group of investigators began to pay attention to these individual cognitive differences. Anchored in laboratory-based experimental psychology, but with a sympathetic stance toward the field of personality, these individuals considered the possibility of extending the then current noncognitive personality field to incorporate the study of relationships between cognition and personality (see Kagan and Kogan, 1970, for an earlier review of this history, and Cantor and Kihlstrom, 1987, for a more recent retrospective review). At that time the personality connection was heavily psychodynamic, a reflection in part of the clinical research settings in which the early pioneering work was carried out.

The psychological traditions out of which cognitive styles emerged were decidedly different from the psychometric and educational traditions that gave birth to the study of the intellectual abilities. For researchers of cognitive style, the psychometric-abilities tradition appeared too encapsulated (because its major concern did not extend beyond the factorial composition of abilities) and too atheoretical (because of the emphasis on prediction of school achievement). By contrast, H. A. Witkin and his collaborators treated the cognitive style of field dependence-independence (FDI) as a master construct influencing motivation, personality, socialization, interpersonal behavior, and even neurophysiological functioning. For the Menninger Foundation group, cognitive styles (labeled "cognitive controls") were located in the "conflict-free sphere of the ego," and a primary goal involved the connections between these cognitive controls and the traditional mechanisms of defense (e.g., Gardner & Moriarty,

1968). This interest in exploring personality and behavioral implications has characterized much of the research in the cognitive-style tradition.

Some description of the constructs that are grouped under the cognitive-style label is clearly warranted. A substantial portion of the cognitive-style field is now almost exclusively of historical interest, so a simple descriptive listing is not useful. Traditions of theory and research that flourished from the late 1940s into the 1970s (and some even into the 1980s) have either disappeared from the psychological scene or have lost much of their vitality. Accordingly, considerably more attention will be paid to those styles that continue to generate research than to those that have remained dormant for an extended period of time.

CLASSIFICATION OF COGNITIVE STYLES

The reader seeking the most comprehensive listing of cognitive styles should consult Messick (1976). Nineteen styles have been delineated there; a definition of each style is provided, and reference is made to the major findings associated with each. A more detailed treatment of some of these styles can be found in Kogan and Saarni (1990).

Because of the great diversity among the various cognitive styles, Kogan (1973, 1983) proposed a three-fold classification scheme based on the nature of the measurement operations employed in the assessment of a style. For Type I styles, individuals can be characterized on the basis of the efficiency (or veridicality) of their performance on criterial tasks; Field Dependence-Independence (FDI) offers a good example of such a style, because both of its primary assessment devices—the Embedded Figures Test (EFT) and the Rod-and-Frame Test (RFT)—generate performance distributions of a better-versus-poorer character (see following section). The original EFT measures the amount of time required to locate designated simple figures embedded in complex geometric designs. In a group-administered version of the EFT, the number of items solved in a prescribed period of time becomes the principal score. For the RFT, whether in its original or its portable form (Oltman, 1968), the respondent's task is to set a vertical rod to the true vertical. A frame encloses the rod, and the frame is systemati-

cally tilted to the right or to the left. Thus, in the EFT, the respondent must overcome the embedded geometric context in order to do well, whereas in the RFT, the respondent must disembed the rod from the surrounding tilt of the frame in order to approximate true verticality. Both of the aforementioned tasks share with ability tests the property of generating individual differences in level or quality of performance. Hence, Type I styles cannot be distinguished from ability on the basis of measurement operations alone.

Another major style with Type I properties—reflection-impulsivity (R-I)—can be traced to the work of Kagan and his associates (1964). The principal measuring instrument for assessing this style is the Matching Familiar Figures Test (MFFT). The MFFT is a perceptual matching task, in which the respondent is required to select one of a set of figural variants (minimally different from each other) that is identical to a standard figure. As originally conceptualized, the MFFT was designed to measure the extent of delay under conditions of response uncertainty, and, in fact, a delay or latency index is obtained. If R-I had relied exclusively on this index, it would obviously not have qualified as a Type I cognitive style. However, it should be noted that Kagan and associates (1964) added a second index to the assessment of R-I, namely the number of errors committed in the selection of variants. Initially, research with the MFFT used a median-split procedure for classifying subjects: long-latency, few-error subjects classified as reflectives and short-latency, many-error subjects classified as impulsives. This procedure in effect excluded the minority of fast accurates and slow inaccurates, and prompted the shift to the use of multiple-regression procedures (Ault, Mitchell, & Hartman, 1976). Of primary interest in the present context is the addition of an accuracy component to the R-I style, hence placing it in the Type I category.

In contrast to Type I styles, those classified as Type II do not employ accuracy or efficiency criteria in their measurement operations, yet performance at one pole of the style is endowed with greater value than performance at a contrasting pole (cognitive complexity-simplicity). The cognitive complexity-simplicity dimension assesses the extent to which people and other aspects of one's social environment are construed in a multidimensional and discriminating manner (e.g.,

Bieri, 1966). A variety of procedures have been employed to assess cognitive complexity-simplicity. All rely on comparisons between target people or objects, with the intention of eliciting the number of independent dimensions the respondent uses to characterize the domain at issue. There is a tendency to value human diversity highly, for example, and hence to believe that a cognitively complex individual relative to a cognitively simple individual would be better able to capture that diversity. Such accuracy cannot be assumed, however, for there are indications that cognitively complex individuals may overattribute differences in the presence of similarity. As Bieri and his associates (1966) have noted, there is an interplay between the complexity of the person and of the situation. Nevertheless, those authors contend that cognitive complexity-simplicity is a pervasive style with a fair degree of generality. Accordingly, positive correlations can be anticipated between complexity and IQ or other ability indices. On the other hand, it has been argued that the present style is completely domain specific—for example, complexity with respect to people independent of complexity with respect to things. If so, one can no longer treat complexity-simplicity as a style, and the question of a style-ability linkage becomes moot.

Styles of conceptualization offer a further example in the Type II category. Object-sorting tasks generate different kinds of groupings. Kagan, Moss, and Sigel (1963) distinguished between thematic and analytic bases of classification. Thematic styles were more frequently produced by younger children, the analytic styles by somewhat older children. Such groupings do not involve matters of accuracy, but the analytic style is presumed to be associated with greater developmental maturity and hence is more highly developed. For any particular chronological age group, therefore, greater preference for the more developmentally mature classification would be associated with higher IQs. In sum, implicit in the Type II category of cognitive styles is the notion of a value distinction implying a more- versus less-advanced level of cognitive functioning. Because matters of correctness or error are irrelevant to Type II styles, however, any link to ability constructs would have to be put to an empirical test. This is especially critical given theoretical disagreement about the developmental maturity of different

styles of conceptualization (see Kogan, 1976). More recently, it has been argued that such styles show no intratask consistency, and that the critical index (likely to be ability related) is flexibility in mode of grouping, that is, showing greater diversity in stylistic preferences.

Finally, Type III cognitive styles are most purely stylistic because their measurement entails neither an accuracy nor an efficiency criterion, and both poles of the style, as originally conceived, are equally valued modes of cognitive functioning. As empirical research on a Type III style proceeds, an array of findings may call into question the original concept of strict value neutrality. In this regard, consider the breadth-of-categorization style—the preferred band width of specified categories. Pettigrew (1958) devised a set of items, each providing the average value of a category (e.g., a normal rainfall in New York), that required the respondent to choose from multiple options the top and bottom extreme of the category (e.g., most and least rainfall in any single year). Verbal and figural bandwidth tasks have also been developed (see Kogan, 1971). At the broad extreme, respondents prefer overinclusiveness, hence stretching category boundaries as far as possible. At the narrow extreme, respondents prefer overexclusiveness, thereby insuring that the category does not include any “inappropriate” instances.

A review by Pettigrew (1982) of almost twenty-five years of research with his bandwidth measure reports an array of correlates (including diverse ability measures). Of primary interest is the evidence that bandwidth preferences do not relate to global IQ but do enter into significant relations—some favoring breadth, others favoring narrowness—with more specific cognitive abilities. Thus, narrow categorizers performed at a higher level on a diversity of memory tasks (Johnson, 1974), whereas broad categorizers were superior in divergent-thinking tasks (Wallach & Kogan, 1965) and in metaphoric processing (Kogan et al., 1980). These outcomes suggest that Type III cognitive styles are not independent of ability; rather, the opposite poles of this style are associated with different kinds of abilities.

Various interpretations have been offered to explain the pattern of ability correlates just described (Pettigrew, 1982). One such interpretation concerns the detailed analytic processing that narrow categorizers

favor in comparison to the allegedly holistic processing favored by broad categorizers. Another concern is the possibly greater tolerance for deviant instances typical of broad categorizers relative to narrow categorizers. Research by Block and his associates (1981, 1986) has demonstrated changes in the meaning of categorization breadth from early to middle childhood. As described by Kogan (1987, p. 112), "an optimal categorization strategy at age 4 ceases to be optimal at age 11. Narrow categorizing serves an adaptive function at the young age, where accurate or rigorous standards of similarity implies a type of selective tuning to environmental requirements. By age 11, a different heuristic assumes prominence, a more adventurous one which allows for the accommodation of deviant but nevertheless appropriate instances." The breadth-of-categorization style is very likely unique in the extent to which it demonstrates so radical a developmental transformation.

To sum up, cognitive styles can be classified into three types depending on the nature of their relationship to the domain of ability. Because Type I styles employ measurement procedures that place a premium on accuracy of performance, the distinction between such styles and specific abilities is often blurred. Type II styles entail value judgments about the quality of response. There is an implicit connection to the domain of ability here (greater developmental maturity implying greater cognitive capacity), but such style-ability links must be established empirically. There has been considerable controversy as to the developmental maturity of various conceptual styles, for example, and value expectations have not always been confirmed. Finally, Type III styles offer no clear guidelines for judgments about accuracy or quality of response on the basis of the style's measurement operations. Rather, each pole of the style in its original formulation simply reflects alternative preferential modes. As we have seen, however, each pole may be linked to distinctively different abilities.

FIELD DEPENDENCE-INDEPENDENCE (FDI)

Given its special role as the historically first and most extensively studied cognitive style, it is only fitting that we devote a special section to FDI. The pri-

mary objective here is to select one aspect of the huge FDI corpus—the style-ability distinction—and to focus upon it.

Witkin and his associates (1962) were well aware of the overlap of FDI with the assessment of intelligence. Those authors, as well as Goodenough and Karp (1961), emphasized the link between FDI and a particular type of intelligence, namely, an analytic factor based on three of the WECHSLER SCALES OF INTELLIGENCE from the Performance battery—block design, object assembly, and picture completion. A common cognitive process—overcoming embeddedness—was presumed to underlie these subtests and the FDI indices derived from the EFT and RFT measures. From these observations, the inference was drawn that the three Wechsler subtests could in fact be assimilated into the FDI construct. In due course, Witkin and Goodenough (1981) subsumed both the EFT and the Wechsler tests within a broad construct of restructuring skill. Consistent with this perspective, one now finds Block Design, for example, employed as an index of FDI.

Just as it is possible to incorporate selected intelligence measures within an FDI construct, so can the converse argument be advanced that FDI cannot be effectively distinguished from traditional indices of intelligence. This latter argument has much in its favor because theories of intelligence have historical precedence over the more recently formulated theory of FDI. Supportive of the view that FDI cannot be psychometrically separated from intelligence are a series of studies demonstrating that FDI indices correlate no more highly among themselves than they do with indices of spatial ability and fluid intelligence. Other investigators have claimed that FDI is highly loaded on the GENERAL INTELLIGENCE (*g*) factor.

The foregoing evidence strongly points to the conclusion that FDI should be situated within the ability rather than the style domain. In an effort to rescue the stylelike status of FDI, Witkin, Goodenough, and Oltman (1979) offered a major reformulation of FDI theory that allowed for distinctive competencies at both the FI and FD poles. Whereas FI individuals were expected to excel in cognitive tasks requiring restructuring, FD individuals were endowed with a diversity of interpersonal skills. In the case of FD, however, it has proven difficult to demonstrate the presence of actual

skill rather than a mere orientation or attentiveness to the social environment (see Davis, 1991; Kogan & Block, 1991). Accordingly, the value-neutral theoretical formulation that Witkin and his colleagues sought to achieve has unfortunately not withstood the test of empirical confirmation.

The style-versus-ability issue is of particular concern to those who approach the FDI topic from a psychometric construct-validated perspective. FDI, however, has also proven to be of interest to investigators working within a neo-Piagetian tradition. Within that tradition, FDI is treated as a performance rather than a competence variable. Casting the FDI construct into a competence-performance framework leads to a distinctly different implication than viewing it as more abilitylike than stylelike from a psychometric perspective. In the competence-performance case, emphasis is given to the information-processing strategies of FI and FD children (see Davis & Cochran, 1990), and to the training interventions required to reduce the FD handicap.

The FDI construct has been distinguished by its enormous breadth of application. There are very few areas of psychology that have remained completely untouched by it. All this could not have come about without the great imagination and energy of the theory's originator, H. A. Witkin. The diversity of FDI correlates would probably have been uncovered if spatial- or fluid-ability measures had been substituted for FDI. The theory of the 1940s underlying the abilities, however, would never have led to the extensions into personality, social psychology, psychotherapy, and neuropsychology that have characterized the work on FDI.

Theory and research on FDI continues as a lively intellectual enterprise, as reflected, for example, by the volumes of Globerson and Zelniker (1989) and Wapner and Demick (1991). Despite a half-century of FDI research, however, we seem to have as many questions as we have answers about the role of FDI in mind and behavior. Matters taken for granted have been reopened to further scrutiny. The long-presumed coherence of FDI (the relationship of EFT and RFT) was called into question in Witkin's final posthumous publication (Witkin & Goodenough, 1981). EFT is now linked to other restructuring tasks, and RFT is presumed to involve the interaction of bodily and visual-

field cues in the perception of the upright. Research on the style-versus-ability issue has already begun to take account of this reformulation of FDI theory (see Kogan, 1983, for some early directions). There is little doubt that the FDI tradition of research has peaked, but this master construct continues to inspire and sustain numerous investigators.

The cognitive-style movement began at a time when the only existing model of intelligence was a psychometric one. With the growing prominence of Piagetian and subsequent neo-Piagetian theories, cognitive styles were thought of as moderators of competence (e.g., Brodzinsky, 1985; Kogan, 1985). The development of newer models of multiple intelligence (Fodor, 1983; Gardner, 1983; Sternberg, 1985) simply came too late to have much of an impact upon cognitive-style theory and research. There is no doubt whatsoever that there are numerous points of connection between these older and newer traditions. Whether such connections will ever be made is one of those unknowns for which only the passage of time can provide an answer.

(See also: MULTIPLE INTELLIGENCES, THEORY OF; PIAGETIAN THEORY OF DEVELOPMENT.)

BIBLIOGRAPHY

- AULT, R. L., MITCHELL, C., & HARTMANN, D. P. (1976). Some methodological problems in reflection-impulsivity research. *Child Development*, 47, 227-231.
- BIERI, J. (1966). Cognitive complexity and personality development. In O. J. Harvey (Ed.), *Experience, structure, and adaptability* (pp. 13-38). New York: Springer.
- BIERI, J., ATKINS, A. L., BRIAR, S., LEAMAN, R. L., MILLER, H., & TRIPODI, T. (1966). *Clinical and social judgment: The discrimination of behavioral information*. New York: Wiley.
- BLOCK, J., BLOCK, J. H., & HARRINGTON, D. M. (1974). Some misgivings about the Matching Familiar Figures Test as a measure of reflection-impulsivity. *Developmental Psychology*, 10, 611-632.
- BLOCK, J., BUSS, D. M., BLOCK, J. H., & GJERDE, P. F. (1981). The cognitive style of breadth of categorization: Longitudinal consistency of personality correlates. *Journal of Personality and Social Psychology*, 40, 770-779.

- BLOCK, J., GJERDE, P. F., & BLOCK, J. H. (1986a). More misgivings about the Matching Familiar Figures Test as a measure of reflection-impulsivity: Absence of construct validity in preadolescence. *Developmental Psychology*, 22, 820–831.
- BLOCK, J., GJERDE, P. F., & BLOCK, J. H. (1986b). Continuity and transformation in the psychological meaning of categorization breadth. *Developmental Psychology*, 22, 832–840.
- BRODZINSKY, D. M. (1985). On the relationship between cognitive styles and cognitive structures. In E. D. Neimark, R. DeLisi, & J. L. Newman (Eds.), *Moderators of competence* (pp. 147–174). Hillsdale, NJ: Erlbaum.
- CANTOR, N., & KIHLMSTROM, J. F. (1987). *Personality and social intelligence*. Englewood Cliffs, NJ: Prentice-Hall.
- CRONBACH, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671–684.
- DAVIS, J. K. (1991). Educational implications of field-dependence-independence. In S. Wapner & J. Demick (Eds.), *Field dependence-independence: Cognitive style across the life span* (pp. 149–175). Hillsdale, NJ: Erlbaum.
- DAVIS, J. K., & COCHRAN, K. F. (1990). An information-processing view of field dependence-independence. In O. N. Saracho (Ed.), *Cognitive style and early education* (pp. 61–78). New York: Gordon & Breach.
- FODOR, J. A. (1983). *The modularity of mind*. Cambridge, MA: M.I.T. Press.
- GARDNER, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- GARDNER, R. W., HOLZMAN, P. S., KLEIN, G. S., LINTON, H. B., & SPENCE, D. P. (1959). Cognitive control: A study of individual consistencies in cognitive behavior. *Psychological Issues*, 1, No. 4.
- GARDNER, R. W., & MORIARTY, A. (1968). *Personality development at preadolescence: Exploration of structure formation*. Seattle: University of Washington Press.
- GLOBERSON, T. (1989). What is the relationship between cognitive style and cognitive development? In T. Globerson & T. Zilniker (Eds.), *Cognitive style and cognitive development* (pp. 71–85). Norwood, NJ: Ablex.
- GLOBERSON, T., WEINSTEIN, E., & SHARABANY, R. (1985). Teasing out cognitive development from cognitive style: A training study. *Developmental Psychology*, 4, 682–691.
- GLOBERSON, T., & ZELNIKER, T. (EDS.), (1989). *Cognitive style and cognitive development*. Norwood, NJ: Ablex.
- GOODENOUGH, D. R., & KARP, S. A. (1961). Field dependence and intellectual functioning. *Journal of Abnormal and Social Psychology*, 63, 241–246.
- JOHNSON, J. H. (1974). Memory and personality: An information-processing approach. *Journal of Research in Personality*, 8, 1–32.
- KAGAN, J., & KOGAN, N. (1970). Individual variation in cognitive processes. In P. Mussen (Ed.), *Carmichael's manual of child psychology* (Vol. 1, pp. 1273–1365). New York: Wiley.
- KAGAN, J., MOSS, H. A., & SIGEL, I. E. (1963). Psychological significance of styles of conceptualization. In J. C. Wright & J. Kagan (Eds.), *Basic cognitive processes in children. Monographs of the Society for Research in Child Development*, 28 (2, Serial No. 86).
- KAGAN, J., ROSMAN, B. L., DAY, D., ALBERT, J., & PHILLIPS, W. (1964). Information processing in the child: Significance of analytic and reflective attitudes. *Psychological Monographs*, 78 (1, Whole No. 578).
- KOGAN, N. (1971). Educational implications of cognitive styles. In G. S. Lesser (Ed.), *Psychology and educational practice* (pp. 242–292). Glenview, IL: Scott, Foresman.
- KOGAN, N. (1973). Creativity and cognitive style: A life-span perspective. In P. Baltes & K. W. Schaie (Eds.), *Life span developmental psychology: Personality and socialization* (pp. 145–178). New York: Academic Press.
- KOGAN, N. (1976). *Cognitive styles in infancy and early childhood*. Hillsdale, NJ: Erlbaum.
- KOGAN, N. (1983). Stylistic variation in children and adolescents: Creativity, metaphor, and cognitive styles. In J. Flavell & E. Markman (Eds.), *Handbook of child psychology: Vol. 3. Cognitive development* (pp. 630–706). New York: Wiley.
- KOGAN, N. (1985). Cognitive styles as moderators of competence: A commentary. In E. D. Neimark, R. DeLisi, & J. L. Newman (Eds.), *Moderators of competence* (pp. 175–189). Hillsdale, NJ: Erlbaum.
- KOGAN, N. (1987). Some behavioral implications of cognitive styles in childhood. *Early Child Development and Care*, 29, 95–117.
- KOGAN, N. (1989). A stylistic perspective on metaphor and aesthetic sensitivity in children. In T. Globerson & T. Zelniker (Eds.), *Cognitive style and cognitive development* (pp. 192–213). Norwood, NJ: Ablex.
- KOGAN, N., & BLOCK, J. (1991). Field dependence-independence from early childhood through adolescence: Personality and socialization aspects. In S. Wapner & J.

- Demick (Eds.), *Field dependence-independence: Cognitive style across the life span* (pp. 177–207). Hillsdale, NJ: Erlbaum.
- KOGAN, N., CONNOR, K., GROSS, A., & FAVA, D. (1980). Understanding visual metaphor: Developmental and individual differences. *Monographs of the Society for Research in Child Development*, 45(1, Serial No. 183).
- KOGAN, N., & SAARNI, C. (1990). Cognitive style in children: Some evolving trends. In O. N. Saracho (Ed.), *Cognitive style and early education* (pp. 3–31). New York: Gordon & Breach.
- MACLEOD, C. M., JACKSON, R. A., & PALMER, J. (1986). On the relation between spatial ability and field dependence. *Intelligence*, 10, 141–151.
- MCKENNA, F. P. (1984). Measures of field dependence: Cognitive style or cognitive ability? *Journal of Personality and Social Psychology*, 47, 593–603.
- MESSER, S. B. (1976). Reflection-impulsivity: A review. *Psychological Bulletin*, 83, 1026–1053.
- MESSER, S. B., & SCHACHT, T. (1983). A cognitive-dynamic theory of reflection-impulsivity. In J. Masling (Ed.), *Empirical studies of psychoanalytic theory* (Vol. 2, pp. 151–195). Hillsdale, NJ: Erlbaum.
- MESSICK, S. (1976). Personality consistencies in cognition and creativity. In S. Messick & Associates (Eds.), *Individuality in learning* (pp. 4–22). San Francisco: Jossey-Bass.
- NEIMARK, E. D. (1981). Confounding with cognitive style factors: An artifact explanation for the apparent nonuniversal incidence of formal operations. In I. E. Sigel, D. M. Brodzinsky, & R. Golinkoff (Eds.), *Piagetian theory and research: New directions*. (pp. 177–189). Hillsdale, NJ: Erlbaum.
- NICKEL, H., PETZOLD, M., RETTIG, K., & VETTER, J. (1985). Conceptual styles and concept formation in childhood: Preferences and competences. *German Journal of Psychology*, 9, 255–267.
- OLTMAN, P. K. (1968). A portable rod-and-frame apparatus. *Perceptual and Motor Skills*, 26, 503–506.
- PASCUAL-LEONE, J. (1989). An organismic process model of Witkin's field dependence-independence. In T. Globerson & T. Zelniker (Eds.), *Cognitive style and cognitive development* (pp. 36–70). Norwood, NJ: Ablex.
- PETTIGREW, T. F. (1958). The measurement and correlates of category width as a cognitive variable. *Journal of Personality*, 26, 532–544.
- PETTIGREW, T. F. (1982). Cognitive style and social behavior. A review of category width. In L. Wheeler (Ed.), *Review of personality and social psychology* (Vol. 3, pp. 199–224). Beverly Hills, CA: Sage Publications.
- DE RIBAUPIERRE, A. (1984). Cognitive style and operational development. A review of French literature and a new Piagetian reinterpretation. In T. Globerson & T. Zelniker (Eds.), *Cognitive styles and cognitive development* (pp. 86–115). Norwood, NJ: Ablex.
- SALKIND, N. J., & WRIGHT, J. C. (1977). The development of reflection-impulsivity and cognitive efficiency. *Human Development*, 20, 377–387.
- SCOTT, W. A., OSGOOD, D. W., & PETERSON, C. (1979). *Cognitive structure: Theory and measurement of individual differences*. Washington: Winston.
- STERNBERG, R. J. (1985). *Beyond IQ: A triarchic framework for intelligence*. New York: Cambridge University Press.
- VERNON, P. E. (1972). The distinctiveness of field independence. *Journal of Personality*, 40, 366–391.
- WALLACH, M. A., & KOGAN, N. (1965). *Modes of thinking in young children*. New York: Holt, Rinehart and Winston.
- WAPNER, S., & DEMICK, J. (EDS.). (1991). *Field dependence-independence: Cognitive style across the life span*. Hillsdale, NJ: Erlbaum.
- WIDIGER, T. A., KNUDSON, R. M., & RORER, L. G. (1980). Convergent and discriminant validity of measures of cognitive styles and abilities. *Journal of Personality and Social Psychology*, 39, 116–129.
- WITKIN, H. A., DYK, R. B., FATERSON, H. F., GOODENOUGH, D. R., & KARP, S. A. (1962). *Psychological differentiation*. New York: Wiley.
- WITKIN, H. A., & GOODENOUGH, D. R. (1977). Field dependence and interpersonal behavior. *Psychological Bulletin*, 84, 661–689.
- WITKIN, H. A., & GOODENOUGH, D. R. (1981). *Cognitive styles: essence and origins*. New York: International Universities Press.
- WITKIN, H. A., GOODENOUGH, D. R., & OLTMAN, P. K. (1979). Psychological differentiation: Current status. *Journal of Personality and Social Psychology*, 37, 1127–1145.
- WITKIN, H. A., LEWIS, H. B., HERTZMAN, M., MACHOVER, K., MEISSNER, P. B., & WAPNER, S. (1954). *Personality through perception*. New York: Harper.
- ZELNIKER, T., & JEFFREY, W. E. (1976). Reflective and impulsive children. Strategies of information processing underlying differences in problem solving. *Monographs of the Society for Research in Child Development*, 41 (5, Serial No. 168).

ZELNIKER, T., & JEFFREY, W. E. (1979). Attention and cognitive style in children. In G. A. Hale & M. Lewis (Eds.), *Attention and cognitive development* (pp. 275–296). New York: Plenum.

NATHAN KOGAN

COHORT EFFECTS The term *cohort* derives from a fundamental organizational unit of the Roman army. In social science *cohorts* are typically defined as groups of individuals experiencing common environmental circumstances (Rosow, 1978; Ryder, 1965). Examples of cohorts are the freshman class at a university, the U. S. Marines who served in combat during the Vietnam war, or all persons exposed to a toxic chemical spill. In demography, it is common to analyze population trends using birth cohorts—members of defined populations or subpopulations born in a particular calendar year (for example, all citizens born in Mexico in 1929).

Cohort effects in intelligence have come to be associated with systematic differences in abilities that are a function of membership in different birth cohorts. Cohort effects are also sometimes referred to as *generational differences*. Differences in the historical timing of birth cause variation in variables that potentially influence intelligence. For example, a person born in 1900 experienced a fundamentally different societal context than did a person born in 1950, given historical changes that have occurred during the twentieth century. In contrast to an American born in 1950, an American born in 1900 was, on average, educated in a different kind of grade school, with a different kind of curriculum; was less likely to attend college; lived on average in a more rural environment; received a quite different quality of medical treatment during childhood; and experienced a vastly different set of specific historical events during maturation and early adulthood (e.g., the former experienced the Great Depression of the 1930s).

Cohort effects are distinct from two other types of related effects: *age changes* and *period effects*. *Age changes* are caused by influences that systematically covary with chronological age, especially developmental processes (e.g., biological aging during adulthood). *Period*

effects refer to influences that affect all birth cohorts alive during a particular historical period. An example of period effects would be public opinion about the likelihood of nuclear war with Russia. One way to understand cohort effects is to view them as the lasting, cumulative effects of events occurring within a limited historical epoch. To illustrate, imagine that cohort effects in intelligence will be caused by an outbreak of a virulent type of viral encephalitis occurring in the year 2000 and ending in 2005. Severe encephalitic trauma can cause brain damage and affect intelligence. If the probability of permanent brain damage was more likely in young children, then the average intelligence of a specific range of birth cohorts (say, persons born between 1995 and 2005) would be lower than that of later-born cohorts, who would not have experienced the epidemic. Earlier cohorts would have been too old at the time of the outbreak to have been affected. The end result would be reduced intelligence in the 1995–2005 birth cohorts, relative to other birth cohorts in the population. Of course, in the real world, there would be a number of potential historical influences on intelligence, and cohort effects across the population would be the outcome of the complex mixture of many such influences.

In the example just given, the cohort is the population of persons affected by the disease. The group is a subpopulation of persons alive between 2000 and 2005 that cuts across multiple years of birth (and hence, multiple birth cohorts). In psychology, cohorts are typically defined by chronological age, perhaps because the historically varying causal variables that define cohorts and that influence constructs like intelligence are not well understood.

Identifying the presence of cohort effects is a difficult task. Part of the problem is due to the fact that there is a systematic relationship between chronological age and membership in a birth cohort. At any given point in historical time, one's age is perfectly predictable from one's year of birth. Thus, at any given point in time, persons of different ages were, by definition, born into different birth cohorts.

Historically, interest in cohort effects in intelligence emerged in developmental psychology because of their importance as a rival explanation for development as a cause of individual differences in intelligence. Devel-

opmental research often assesses age-related changes by employing a cross-sectional design. In this design, persons of different ages are sampled at a single point in time. If age differences in measures of intelligence are observed, then one can infer the presence of age-related changes in intelligence. However, given the linear dependency between year of birth, current chronological age, and historical time, a plausible rival explanation for the age differences is the existence of cohort effects. Flynn (1987) conducted an analysis of standardization data on the Wechsler Adult Intelligence Scale collected over many decades, using a time-lag analysis. He argued that there have been "massive" gains in intelligence test scores during the twentieth century, with more recently born birth cohorts performing at higher levels. This pattern of effects would cause older persons to perform more poorly than younger persons, producing age differences in cross-sectional data. The differences could not reasonably be viewed as pure estimates of the effects of biological aging, however. Thus, with respect to the hypothesis of developmental change, cohort effects can be viewed as an unwanted nuisance factor that makes it difficult to interpret the results from a simple cross-sectional design. Given that birth cohort and chronological age are perfectly confounded in a simple cross-sectional design, there is no way to use cross-sectional data to estimate their separate influences.

Schaie (1965) was one of the first to propose using more elaborate sampling strategies to determine (1) if there are cohort effects in intelligence, and (2) whether these cohort effects distort estimates of age changes taken from cross-sectional designs. His approach involved repeated sampling from the same population across multiple time periods. There are two basic sampling schemes: one repeatedly draws new cross-sectional samples at different time periods (making a cross-sectional sequence), and the other retests the participants in the first cross-sectional sample as they grow older (producing a longitudinal sequence). Schaie (1983) ran a unique long-term study that combines both cross-sectional and longitudinal sequences to study multiple primary intellectual abilities (as defined by Thurstone, 1938). Schaie's research project began in 1956 with a cross-sectional sample of adults ranging in age from 21 to 70. Every seven years, all persons still willing to participate were retested and a

new cross-sectional sample was added. Using this large data set, Schaie found evidence of cohort effects in several, but not all, intellectual abilities. He has reported cohort effects for inductive reasoning, spatial relations, verbal comprehension, and numerical facility. The pattern of these cohort effects has been consistent with the hypothesis that traditional cross-sectional designs may overestimate age changes in abilities like inductive reasoning, because earlier-born cohorts do more poorly on the reasoning tests. However, Schaie's data suggest that the largest cohort differences occurred for birth cohorts born from 1890 through about 1950, followed by a leveling off of the trend. In some cases, the trend for increasing performance by new birth cohorts appears to have been reversed. The cohort effects on numerical facility, measured with a test of two-column addition problems, are an interesting example. They suggest that today's younger adults (the most recently born adult birth cohorts) perform less well than earlier birth cohorts (today's middle-aged and older persons), adjusting for age-related trends. This pattern of cohort effects is consistent with recent societal trends: reduced emphasis on rote addition skills, combined with the increased use of hand calculators and computers. Schaie has found no cohort effects on measures of perceptual speed, requiring rapid identification of visual and symbolic information. Results from a different sequential study (Cunningham & Tomer, 1990) also suggest no cohort differences in perceptual speed, while finding cohort effects for other intellectual abilities (see also Alwin, 1990).

Schaie's results and conclusions have generated a fair amount of argument and controversy, however. One problem is that the kind of sequential sampling required to investigate whether there are cohort effects is both expensive and time consuming. As a result, there are very few studies that have relevant data. Like any particular study, Schaie's results could be specific to the types of intelligence tests he used, the methods of sampling persons, and so on. Without additional long-term sequential studies of the same phenomena, we do not know whether cohort effects will be replicated. A second problem is that Schaie and his colleagues have used the finding of significant cohort effects to argue that aging changes occur later in life, are far smaller in magnitude, and are found for some

persons but not others. The interpretation has been highly controversial (see Baltes & Schaie, 1976; Horn & Donaldson, 1976). To some degree, the very interesting questions of how to think about and understand cohort effects have been overshadowed by the debate on their implications for understanding the aging of intelligence.

By far the biggest problem, however, is a technical one. There are inherent ambiguities in the interpretation of statistical estimates of cohort effects in sequential data sets, given the mathematical dependency among the age, period, and cohort variables alluded to above. The sequential sampling plan affords the scientist the chance to estimate age and cohort effects because it uses repeated sampling to give observations of multiple cohorts over the same age ranges. However, it cannot break the mathematical dependency between the three time-related variables. When one has repeatedly sampled multiple birth cohorts over the same age ranges, one has also, by definition, measured the different cohort groups across different historical time periods. In a sense, the issue arises because of the use of birth cohort and chronological age to index the historical and developmental causes of intelligence. That is, although age, cohort, and time-period effects are produced by conceptually distinct phenomena, no statistical analysis can estimate all the conceivable effects using chronological age, year of birth, and year of testing as the index variables for these effects. Instead, one needs to make some assumptions, based upon theory, that at least some of the age, period, and cohort effects are not operating. Schaie's statistical methods for estimating cohort effects have been severely criticized because they often make very strong assumptions—such as a complete absence of period effects (e.g., Adam, 1978). There is little question that the type of assumptions one makes about the presence or absence of particular age, period, and cohort effects can greatly influence the statistical estimates of the remaining effects (Donaldson & Horn, 1992; Glenn, 1977; Schaie & Hertzog, 1982). Given that psychologists differ widely in their beliefs about the nature and causes of intelligence, this topic will likely continue as a source of debate and controversy.

(See also: SOCIOECONOMIC STATUS AND INTELLIGENCE.)

BIBLIOGRAPHY

- ADAM, J. (1978). Sequential strategies and the separation of age, cohort, and time-of-measurement contributions to developmental data. *Psychological Bulletin*, 85, 1309–1316.
- ALWIN, D. F. (1990). Family of origin and cohort differences in verbal intelligence. *American Sociological Review*, 56, 625–638.
- BALTES, P. B., & SCHAIE, K. W. (1976). On the plasticity of intelligence in adulthood and old age. *American Psychologist*, 31, 720–725.
- CUNNINGHAM, W. R., & TOMER, A. (1990). Intellectual abilities and age: Concepts, theories, and analyses. In E. A. Lovelace (Ed.), *Aging and cognition: Mental processes, self-awareness, and interventions* (pp. 379–406). Amsterdam: North-Holland.
- DONALDSON, G., & HORN, J. L. (1992). Age, cohort, and time developmental muddles: Easy in practice, hard in theory. *Experimental Aging Research*, 18, 213–222.
- FLYNN, J. R. (1987). Massive IQ gains in fourteen nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171–191.
- GLENN, N. D. (1977). *Cohort analysis*. Beverly Hills, CA: Sage Publications.
- HORN, J. L., & DONALDSON, G. (1976). On the myth of intellectual decline in adulthood. *American Psychologist*, 31, 701–719.
- ROSOW, I. (1978). What is a cohort and why. *Human Development*, 21, 65–75.
- RYDER, N. B. (1965). The cohort as a concept in the study of social change. *American Sociological Review*, 30, 843–861.
- SCHAIE, K. W. (1965). A general model for the study of developmental problems. *Psychological Bulletin*, 64, 92–107.
- SCHAIE, K. W. (1983). The Seattle longitudinal study: A 21-year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 64–135). New York: Guilford Press.
- SCHAIE, K. W., & HERTZOG, C. (1982). Longitudinal methods. In B. B. Wolman (Ed.), *Handbook of developmental psychology* (pp. 91–115). Englewood Cliffs, NJ: Prentice-Hall.
- THURSTONE, L. L. (1938). Primary mental abilities. *Psychometric Monographs* (No. 1). Chicago: University of Chicago Press.

CHRISTOPHER HERTZOG

COLLEGE BOARD The College Board, once known as the College Entrance Examination Board, was created at the turn of the twentieth century to bring order out of chaos in college admissions. Previously, each college had administered its own examinations to applicants. Members of the faculty wrote the examinations, sometimes in haste, which were administered to juniors and seniors in secondary schools, who frequently journeyed to the campuses on a Saturday to sit for them. Not only was the whole testing apparatus clumsy and the faculty-made examinations somewhat erratic, the examination procedure began all over again if a student decided that he or she preferred another college.

One of the overriding concerns was the level of excellence expected of students. A secondary concern was curriculum content, the preparation that the secondary schools were expected to provide. On both counts, especially the latter, marked differences existed from college to college. For example, at Columbia University in 1878, entrance examinations covered the four books of Caesar's *Commentaries*, Sallust's *History*, Xenophon's *Anabasis*, Greek and Latin prose composition, and ancient geography. The Harvard examination patterns differed substantially. In 1880, Harvard reduced the amount of required Greek and Latin and introduced eight new subjects: algebra, geometry, physical geography, English grammar, English composition, ancient history, and United States history. Not only did the views of various colleges differ, but they were strongly held. Organization of the examination process was an obvious necessity for fairness to applicants, but it was difficult to achieve.

To meet this need, Nicholas Murray Butler, who later became an outstanding president of Columbia, convened a group of college representatives to consider a proposal to establish what would become the College Board.

If there was any single act that brought the College Board into existence, however, it was the challenge posed by Ethelbert Warfield, President of Lafayette College, who asserted that his institution was not going to be told whom to admit. President Eliot, of Harvard, rose and replied with exquisite irony that the proposals had been misunderstood. President Warfield, if he so chose, could admit only those students who were unable to pass these examinations; no one

proposed to deprive LaFayette College of that privilege. The assembled group then voted to form the College Board.

Now that some 2.5 million students annually take the SCHOLASTIC ASSESSMENT TEST (SAT), achievement tests, and the Preliminary Scholastic Aptitude Test/National Merit Scholastic Qualifying Test (PSAT/NMSQT, a shortened version of the SAT for juniors), the small beginnings of the board examinations are difficult to imagine. In 1901, at 67 centers in America and two in Europe, only 973 candidates sat for examinations. Thirty-nine men and women sat around a table in the Columbia library grading the 7,889 papers these 973 candidates wrote. The primary function of the board during the next half-century was to administer and score examinations. It operated on small numbers and generally ran a deficit.

During these years there were some amusing answers to the examination questions, although perhaps none better than one to the question, Name two ancient sports. The classic answer: "Antony and Cleopatra."

Group intelligence tests, forerunners of the SAT, were administered to army recruits during World War I because the results were helpful for placement and training purposes. Carl C. BRIGHAM, a member of the team that created these tests for the military, took this idea back to Princeton University after the war and, with some modifications, began to use general ability tests as an aid in making decisions about applicants applying for scholarships. Research studies indicated that this test might also prove to be a useful addition to the subject-matter essay tests offered by the College Board. In 1926, the first multiple-choice SAT was administered to some 8,000 candidates. Three years later, the test featured verbal and mathematical sections. In 1941, a 200-to-800-point scale was established. Before that time norms were set on whichever group happened to sit for a particular administration of the test.

Until the SAT scores became available, the admissions process in American colleges was based heavily on lineage and proper schooling. The decisions even revealed some unfortunate religious and racial overtones. Now much of this process has changed. Decisions are seldom made on the basis of birth, schooling, and geography and much more on merit, which has improved the intellectual strength of student bodies,

especially at selective colleges. The resulting diversity of student bodies is probably more dramatic than the change in composition of any other institution in our society.

Has the SAT been a fundamental ingredient in this revolution? The answer is yes, in a variety of ways. The SAT should receive credit for helping to eliminate religious influences. It is embarrassing even to suggest that a Jewish applicant who had SAT scores totaling 1550 but applied from an unknown school was not as strong as another applicant from a well-known school whose scores totaled only 1100, assuming both had many As on their grade record. The SAT reduced differences between known and unknown secondary schools because scores validated grades. The SAT has also accelerated the admission of Asians who scored well but whose parents were poorly educated; and although it did not facilitate the entry of African-American students to predominantly white colleges, when coupled with affirmative action programs it has not been the handicap many expected it to be. Of course, most African-American students in no way need the help suggested by affirmative-action programs. They are admitted for their own obvious strengths, as indicated by the number enrolled in virtually every selective institution.

To improve efficiencies in the administration of an increasing volume of tests and to strengthen the research associated with test development, in 1947 a new organization, the Educational Testing Service (ETS), was created by the College Board, the American Council on Education, and the Carnegie Foundation for the Advancement of Teaching. At that time ETS assumed the responsibility for the administration and development of the SAT for the College Board. To many students, it is ETS, not the College Board, that appears to sponsor the SAT, the achievement tests, and the advanced placement exams. Since its founding, ETS has conducted research in a variety of assessment areas and created a number of graduate and professional examinations, in addition to those for the College Board.

Over the years the role of the College Board has continued to evolve and expand. Long identified solely with the SAT and achievement tests, it became a strong membership organization representing the views of schools, colleges, and students nationally and offer-

ing a variety of other services to meet the needs it identified. For example, the cost of higher education is now so large that financial assistance has become the keystone to college attendance for most students. The College Board, again with ETS assistance, has offered through the College Scholarship Service (CSS) a method of processing financial aid information known as the Financial Aid Form (FAF), which is completed by students and parents. It contains information essential in determining the size of need-based awards. Competing with the FAF is a less expensive, albeit less effective, form available from the federal government.

Why has the College Board, a nonprofit, membership organization of schools and colleges dedicated to facilitating the transition of students from school to college, sometimes attracted criticism? Virtually all the charges have arisen from the board's sponsorship of the SAT, which some people claim is biased against certain ethnic minorities and against females, subject to score increases by coaching, and influenced by parental income. Most studies addressing these charges find that they have little validity. SAT scores do improve the prediction of college grades and, when combined with the secondary school record, are useful in admissions. Tutoring schools, which appear to be effective because their students score higher on the second try, generally attract those who originally scored poorly, so that regression toward the mean alone improves their scores. Even so, the score increases generally fall within the standard error of the test, some 30 to 35 points, which, on a scale of 200–800, is quite marginal. Members of some ethnic groups who do not receive the best secondary education find that their lower SAT scores reflect their lack of educational opportunities. They also tend to receive grades in college in line with these scores. In quite a different way, this situation is also true for females. The basic claim is that because women receive higher grades than men in both high school and college but lower SAT scores, the test is biased. First, the differences between female and male grades and scores are small. Second, women enroll in more humanities courses, in which the grades are higher, than do men. If an adjustment is made for those grade differences, the apparent discrepancy disappears. Finally, a small but significant correlation exists between family income and SAT scores; people

with more material resources do provide their children with a better education.

Although the SAT is far from a perfect instrument, when used with other information about the candidates it provides information that is useful for making admissions decisions and for course placement. The critics fail to recognize how much the SAT has improved both the quality and the equity of admissions decisions, and how influential it has been in moving the process toward evaluation based on merit rather than on privilege. Another important value of the scores is that they determine to which colleges students apply more than grades do. A student in a rural school or even an urban school who scores very well will say "I'm good enough to go to one of those fancy colleges." A student seeks admission to the most selective institutions because of scores and not because she or he was valedictorian. (Every school has one.) Furthermore, despite good grades, a student does not apply to MIT or Cal Tech with a score of 500 on the math part of the SAT. Such an option is much more feasible if the student's score is 750. The SAT facilitates the great sorting process, which in turn reduces the range of scores at any one college. This reduction lowers the correlation between SAT scores and college grades. The test then seems less effective than it actually is. That the useful variance of the test is "used" in the admission process, however, is just what should happen.

The tests also serve another function. It is becoming more and more difficult for a college to differentiate among students by using secondary school grades alone, because while grade inflation is pronounced in the colleges, it is even worse in high schools. One school, for example, had twenty-three valedictorians in a single year—it established a grade average that entitled a student to be classified as first in class and over a hundred achieved this average. Actually, 30 percent of those students who take the SAT report that they have grade averages of A. That situation presents a special challenge to the college admissions process; the SAT scores provide an important part of a student's dossier to start the annual sorting process.

The board now provides a wide variety of educational services in addition to the admission test series of SAT I and II. Colleges can use the Admitted Student

Questionnaire to understand their student body better. The Enrollment Planning Service helps colleges organize their recruiting efforts. The PSAT/NMSQT provides practice for the SAT and guidance services for high school juniors and assists in the identification of National Merit Scholars. The Student Search Service helps colleges contact students who may be interested in attending their institutions, and the College Scholarship Service processes important information through the Financial Aid Forms. Advanced Placement examinations, the College Level Examination Program, and Accuplacer all assist in placing students at the proper level in various courses. The College Board also sponsors two educational reform projects—Equity 2000 and Pacesetter—that aid the transition to college of students from diverse backgrounds. Equity 2000 helps secondary school students by eliminating tracking and focusing on mathematical skills and achievement, which research has demonstrated is the gatekeeper to college. Pacesetter has five capstone courses designed to motivate students to high standards of achievement in analytical thinking and levels of mastery. In addition, the board offers a number of associated services (workshops, institutes, regional meetings, and the National Forum) centrally and through the regional offices.

(See also: SCHOLASTIC ASSESSMENT TESTS.)

BIBLIOGRAPHY

- ANGOFF, W. H. (1971). *The College Board admissions testing program*. New York: College Entrance Examination Board.
- COLLEGE ENTRANCE EXAMINATION BOARD. (1926). *The work of the College Entrance Examination Board: 1901–1925*. New York: Ginn.
- COLLEGE BOARD. (1986). *Measures in the college admissions process: A College Board colloquium*. New York: College Entrance Examination Board.
- CROUSE, J., & TRUSHEIM, D. (1988). *The case against the SAT*. Chicago: University of Chicago Press.
- DONLON, T. F. (1984). *The College Board technical handbook for the Scholastic Aptitude Test and Achievement Tests*. New York: College Entrance Examination Board.

- FUESS, C. M. (1950). *The College Board: Its first fifty years*. New York: Columbia Press.
- HANFORD, G. (1990). *Life with the SAT*. New York: College Entrance Examination Board.

DEAN K. WHITLA

COMPENSATORY MECHANISMS The concept of compensation has received considerable attention in contemporary research on intellectual development in adulthood. Two basic issues are (1) the extent to which older adults overcome aging-related decline in intellectual abilities, and (2) the mechanisms through which such compensation may occur.

In the most general sense, compensation is a process through which deficits or losses are moderated through alternative strategies, mechanisms, or selective improvements. Compensation is an important concept in a wide range of psychological literature. Bäckman & Dixon (1992) indicated that compensation occurs systematically in such research areas as sensory handicaps, in which loss in one sensory system may be balanced by increased physiological sensitivity in another intact system; cognitive deficits and learning disabilities, in which alternative methods and investing more effort and time have been explored as potential compensation mechanisms; and neurological (brain) injury, in which both (automatic) biological and (intentional) behavioral compensatory mechanisms have been studied.

The many sensory, neurological, and cognitive changes that occur in adulthood are related to changes in performance on intelligence tests. Researchers interested specifically in adult development have turned increasing attention to the role of compensation in accounting for some successful or adaptive performance in late life. For example, in the social psychology of aging, Kenneth Ferraro (1984) observed that social participation may increase immediately after the death of a spouse, followed by a decrease in level of social involvement, but with a new balance of close social contacts. These new levels of nonmarital intimacy are thought to compensate for the loss of a long-term spouse. In some aspects of cognitive aging a similar substitution may occur. For example, Timothy Salt-

house (1984) observed that older skilled typists, although less proficient on critical components of typing (e.g., finger-tapping rate), are able to maintain high levels of performance by looking further ahead to the to-be-typed material than do younger typists. In the widowhood example a significant personal loss (of a spouse in late life) and in the typing example an apparent decline (in the mechanism of performing a complex task) were counterbalanced by gains in substitutable domains.

Several conceptual treatments of compensation in cognitive aging are available. For Salthouse (1990), clear evidence of a balance between declining components of cognitive skills and improving components is required for an inference of compensation. In contrast, Paul Baltes's (1987) model of selective optimization with compensation portrays the dynamic interaction between improving and declining aspects of intelligence in terms that allow for compensation to occur through increasing specialization. That is, not only may a cognitive skill be performed through different mechanisms (substitution, as with Salthouse) but some cognitive skills may be selectively culled from one's repertoire (and others emphasized) because of declining intellectual resources. According to Baltes, this situation too is an example of compensation and, perhaps, a prototypical example of successful aging.

Compensation is a widely investigated psychological phenomenon, but it may take different forms within and across domains. The definition of Lars Bäckman and Roger Dixon (1992, p. 272) reflects this breadth:

Compensation can be inferred when [a] mismatch between accessible skills and environmental demands is counterbalanced (either automatically or deliberately) by investment of more time or effort (drawing on normal skills), utilization of latent (but normally inactive skills), or acquisition of new skills, so that a change in the behavioral profile occurs [typically] in the direction of adaptive attainment, maintenance, or surpassing of normal levels of proficiency.

Despite this breadth, however, several salient characteristics of compensation emerge. First, a basic assumption is that compensatory behavior occurs only in the context of a *mismatch* (a deficit or loss) between environmental demands and the cognitive skills of an

individual. The mismatch may result from aging-related decline in neurological structures, such as that resulting in decrements in fluid intelligence (Horn, 1978, 1982). The mismatch may also derive from a deficit resulting from an injury or a congenital condition, impairing performance on cognitive tasks. The mismatch may also be the consequence, however, of increasing demands from the environment, such as when cognitive performance does not change, but the level required for successful performance increases.

Second, the *mechanisms* of compensation are critical to consider. The general goal for the compensating individual, of course, is to decrease the gap between the environmental demands and one's performance level. The literature on compensation suggests that there are several ways in which this match can be achieved. Two major forms of compensation are investment of more effort, such as trying harder or taking longer to perform the cognitive task, and using a substitutable skill or component, either of which may be latent or acquired. Other forms of compensating include changing goals, that is, individuals may modify their expectations for performance and criteria for success and select alternative tasks or goals. What is the evidence that these compensatory mechanisms occur in intellectual development?

COMPENSATION IN INTELLECTUAL AGING

For compensation to occur in intellectual development, some degree of plasticity, diversity (individual differences), or metacognitive awareness may be present. Plasticity is important. If some changes in intellectual performance (e.g., normal declines with aging) are not reversible in principle—whether through the same or different mechanisms—then the logic of compensation is at risk. Individual differences in intellectual development are also an important consideration. Universal compensation for intellectual decline would not be expected (Schaie, 1983). Indeed, individuals may be more or less adept at developing compensatory mechanisms. Finally, awareness of a loss or decline may be an important precursor to some compensatory efforts, although research indicates that compensation may also originate in an automatic response and that

it may become automatized with practice (Bäckman & Dixon, 1992).

In a review of the history of research on intellectual development in adulthood, Roger Dixon, Deirdre Kramer, and Paul Baltes (1985) noted that many of these issues—including compensation—occurred in the initial research reports on intellectual aging almost a century ago. For example, E. C. Sanford (1902) linked intellectual decline with the physical decline occurring with aging. Nevertheless, he argued that at least for young-old adults (ages 55–70) continued effort and activity could result in some resistance to inevitable decline. He wrote that “intellectual vigor may survive (and as sometimes happens, much more than compensate the failure on the physical side), but [the individual] must take care of [oneself] . . . that [the] body [must] be able to support the demands of [the] mind. In intellectual matters, even, [people] may find that [they] must fight [their] indolence” (Sanford, 1902, p. 447). Salthouse (1990) quotes several examples of allusions to compensation in intellectual aging. For example, Miles (1942, p. 772; quoted in Salthouse, 1990) wrote that “well-formed and practiced mental habits plus the knowledge increment may tend to compensate in later age for the quickness in comprehension and action that typify early maturity.” Alan Welford (1958, p. 286; quoted in Salthouse, 1990) wrote that “older people have a remarkable ability to compensate for any changes which may tend to impair their performance and show an automatic and unconscious rendering of their activity to make the best use of what capacities they have . . . a process . . . we may call unconscious optimization.”

RESEARCH

Because the preceding sections have established that compensation in intelligence is neither an entirely novel nor inconceivable idea, a brief review of relevant research is in order, particularly studies pertaining to the major forms of compensation in intellectual aging. First, compensation through investing more time or effort could be observed if trying harder resulted in performance improvements. To measure effort at the time of task performance is complicated. Nevertheless, Richard Jennings and colleagues (1990) found that

heart-rate patterns during high-load memory performance indicated an increased expenditure of effort by older adults, an increase that was related to memory performance. Older adults are dramatically affected when tasks, such as those tapping fluid intelligence, require speeded performance (Horn, 1982). Changing the preparation for a fluid intelligence task by providing practice or training changes the nature of the task. Nevertheless, some research has suggested that impaired individuals may improve their performance through training, that is, as a function of effort to acquire and implement strategies (Baltes & Willis, 1982). In a similar vein, through extensive effort older normal adults learned complicated strategies to perform at very high levels in fluidlike tasks such as serial-digit memory (Kliegl & Baltes, 1987). Indeed, with training in a particular mnemonic technique and under self-paced conditions, some older adults remembered as many as 80 digits in the correct serial position. Incorporating knowledge (through mnemonic techniques) into task performance may effectively change the task from fluid to crystallized intelligence. (See FLUID AND CRYSTALLIZED INTELLIGENCE, THEORY OF.) Because of this argument and other reasons, the latter illustrations fall short of being definitive examples of this form of compensation. They do, however, underscore the relevance of investment of effort and time in improving performance and possibly overcoming deficits.

The second general mechanism of compensation is substitution of an alternative ability or component thereof. In psychometric intelligence and aging research, the following scenario is conceivable: With fluid abilities (G_f) declining with advancing age, crystallized abilities (G_c , which may be maintained until later in life) may assume a greater role in accounting for overall intelligent behavior. According to John Horn (1978), such a scenario (or the reverse) may occur because the two dimensions of intelligence are compounds of multiple—sometimes overlapping or simultaneously contributing—cognitive abilities. This situation implies that certain cognitive tasks may be performed through the operation of a typical mechanism (those indexed by G_f) or through an alternative mechanism (those indexed by G_c). Horn's (1978, p. 222) example is instructive: For a given compound task (verbal analogies), relatively high G_c (e.g., vocab-

ulary) may compensate for relatively low G_f (e.g., reasoning). Marit Olofsson and Lars Bäckman (1992) obtained a similar pattern. Whereas reading comprehension was predicted by reading span for younger adults, it was influenced by different markers of preserved semantic memory functioning by older adults.

The compensatory mechanism of substitution has been examined in more detail for tasks that are less psychometric in nature. A particularly useful procedure for examining the interacting role of multiple component processes in the performance of complex cognitive tasks has been produced by Neil Charness (1989) and by Salthouse (1985, 1990). In this procedure—known as the *Molar Equivalence and Molecular Decomposition strategy*—age equivalence on a molar cognitive skill is accounted for either by age equivalence on the typical molecular components of that skill, or by an increase in alternative compensatory molecular components. This procedure is applicable to any ability or skill that can be decomposed into constituent components. In early studies on chess and bridge, evidence suggestive of compensation was found in that, whereas no age differences existed for skill or skill-related tasks, reliable age differences applied to memory tasks relevant to the skill on the part of older experts (Charness & Bosman, 1990). This situation suggests that compensation did occur, but it does not identify the compensatory mechanism. As noted earlier, research on the skill of transcription typing has taken this next step. Although older skilled typists performed the molar task at an equivalent level to younger skilled typists, they performed the molecular components (finger tapping speed, choice reaction time) worse than their younger counterparts (Salthouse, 1984). The compensatory mechanism identified in this research is that of preview span; older adults began keystroke preparation earlier than younger typists thus moderating the role of the highly speeded (and declining) molecular components (Bosman, 1993; Salthouse, 1984).

Finally, another approach to research on compensatory mechanisms in intellectual aging focuses on the substitution of the skills of close collaborators (Dixon, 1992). The basic issue is the extent to which older individual adults who are experiencing cognitive decline can collaborate effectively with others to perform

at a level equivalent or superior to that of individual and collaborating younger adults. This approach has been explored with several cognitive tasks, including those tapping *Gf* and *Gc* (Dixon, 1992).

CONCLUSION

Compensation is a process through which deficits or losses are counterbalanced through alternative strategies, mechanisms, or gains. The life-span profiles of intellectual development are complex and sometimes controversial. Although there is clear evidence for decline on numerous basic cognitive abilities, evidence has emerged for both individual differences in rate and extent of decline, and maintenance in domains of cognitive skill. Both experience (expertise) and compensation may play a role in these phenomena. Two major categories of compensatory mechanisms are investment of more effort and applying a substitutable skill. Although reasonable arguments and data have been proffered regarding both mechanisms, further research will elucidate their operation in intellectual aging.

(See also: LEARNING DISABILITY; MENTAL RETARDATION, CULTURAL-FAMILIAL.)

BIBLIOGRAPHY

- BÄCKMAN, L. (1989). Varieties of memory compensation by older adults in episodic remembering. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 509–544). Cambridge, England: Cambridge University Press.
- BÄCKMAN, L., & DIXON, R. A. (1992). Psychological compensation: A theoretical framework. *Psychological Bulletin*, *112*, 259–283.
- BALTES, P. B. (1987). Theoretical propositions of life-span developmental psychology: On the dynamics between growth and decline. *Developmental Psychology*, *23*, 611–626.
- BALTES, P. B., & WILLIS, S. L. (1982). Plasticity and enhancement of intellectual functioning in adulthood: Penn State's Adult Development and Enrichment Project (ADEPT). In F. I. M. Craik & S. E. Trehub (Eds.), *Aging and cognitive processes* (pp. 353–389). New York: Plenum.
- BOSMAN, E. A. (1993). Age-related differences in the motoric aspects of transcription typing skill. *Psychology and Aging*, *8*, 87–102.
- CHARNESS, N. (1989). Age and expertise: Responding to Taland's challenge. In L. W. Poon, D. C. Rubin, & B. A. Wilson (Eds.), *Everyday cognition in adulthood and late life* (pp. 437–456). Cambridge, England: Cambridge University Press.
- CHARNESS, N., & BOSMAN, E. A. (1990). Expertise and aging: Life in the lab. In T. M. Hess (Ed.), *Aging and cognition: Knowledge organization and utilization* (pp. 343–385). Amsterdam: Elsevier/North-Holland.
- DIXON, R. A. (1992). Contextual approaches to adult intellectual development. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 350–380). Cambridge, England: Cambridge University Press.
- DIXON, R. A. & BÄCKMAN, L. (1993). The concept of compensation in cognitive aging: The case of prose processing in adulthood. *International Journal of Aging and Human Development*, *36*, 199–217.
- DIXON, R. A., KRAMER, D. A., & BALTES, P. B. (1985). Intelligence: A life-span developmental perspective. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 301–350). New York: Wiley.
- FERRARO, K. F. (1984). Widowhood and social participation in later life: Isolation or compensation? *Research on Aging*, *6*, 451–468.
- HORN, J. L. (1978). Human ability systems. In P. B. Baltes (Ed.), *Life-span development and behavior* (Vol. 1) (pp. 211–256). New York: Academic Press.
- HORN, J. L. (1982). The theory of fluid and crystallized intelligence in relation to concepts of cognitive psychology and aging in adulthood. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 237–278). New York: Plenum.
- JENNINGS, J. R., NEBES, R. D., & YOVETICH, N. A. (1990). Aging increases the energetic demands of episodic memory: A cardiovascular analysis. *Journal of Experimental Psychology: General*, *119*, 77–91.
- KLIEGL, R., & BALTES, P. B. (1987). Theory-guided analysis of mechanisms of development and aging through testing-the-limits and research on expertise. In C. Schooler & K. W. Schaie (Eds.), *Cognitive functioning and social structure over the life course* (pp. 95–119). Norwood, NJ: Ablex.
- OLOFSSON, M., & BÄCKMAN, L. (1992). Components of reading comprehension across the adult life span. *European Journal of Gerontology*, *1*, 285–292.

- SALTHOUSE, T. A. (1984). Effects of age and skill in typing. *Journal of Experimental Psychology: General*, 113, 345–371.
- SALTHOUSE, T. A. (1985). *A theory of cognitive aging*. Amsterdam: North-Holland.
- SALTHOUSE, T. A. (1990). Cognitive competence and expertise in aging. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 310–319). San Diego, CA: Academic Press.
- SANFORD, E. C. (1902). Mental growth and decay. *American Journal of Psychology*, 13, 426–449.
- SCHAEIE, K. W. (1990). Intellectual development in adulthood. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (3rd ed., pp. 291–309). San Diego, CA: Academic Press.

ROGER A. DIXON
LARS BÄCKMAN

COMPETENCE VERSUS PERFORMANCE

The competence/performance distinction contrasts what one knows (competence) with how that knowledge is actually used (performance). For example, a fifth-grader may understand the principles of long division perfectly but still fail a math test consisting of long-division problems. The teacher wishes to assess the child's grasp of a mathematical principle, but the test score reflects more than that, including the clarity of the instructions, knowledge of English, familiarity with test-taking conventions, and how tired, distracted, nervous, impulsive, or careless the child is. The competence/performance distinction is controversial on several grounds (see below). However, it is assumed to be important, because behavior is never a perfect reflection of underlying knowledge.

The competence-performance distinction has important theoretical and practical implications. From a theoretical viewpoint, the distinction assumes that people make use of unconscious, rule-governed knowledge, and that such knowledge is not directly reflected in observable behavior. Instead, contrary to behaviorist psychology, outward behaviors can be understood only in terms of the underlying mental states they reveal. Competence and performance often take very different forms from one another. To take the example of long division again, performance consists of the solution to a list of specific problems,

whereas competence consists of general procedures that can be applied to any problem.

From a practical standpoint, the distinction suggests that measuring what people know is a challenging task requiring sensitivity and creativity. Furthermore, any test result (school exams, IQ tests, achievement tests, college boards) must be interpreted cautiously. Every assessment reflects a complex mix of factors, only some of which are of interest to the tester.

BACKGROUND AND CONTROVERSIES

The contrast between competence and performance gained wide attention with the writings of the eminent linguist Noam Chomsky, who argued for a sharp division between outward speech and tacit (non-conscious) linguistic knowledge. Everyday speech contains many errors that are not part of the speaker's knowledge of English, but simply the result of other intervening factors. For example, in casual conversation you may start one thought then abandon it mid-stream, forget the beginning of your sentence, and use the wrong form of the verb. However, if given more time to plan and reflect, you could avoid the errors.

But competence is not simply equivalent to performance minus errors, because what speakers know of their language is more than simply a list of grammatical sentences. People are creative in their use of language; they can speak and understand sentences that have not been uttered before. Moreover, the underlying grammatical rules that comprise one's competence can generate sentences that would never be spoken. For example, languages permit infinitely long sentences that embed one phrase within another ("This is the cat that ate the rat that ate the cheese that . . ."). The fact that the sentences people actually speak are finite is considered an artifact of short-term memory limitations (performance) rather than part of the grammar.

Although the contrast between competence and performance has been a fruitful assumption in linguistics and in other fields, many scholars have criticized the distinction as articulated by Chomsky. One criticism is that the distinction may place too much emphasis on competence, framing it as more interesting and more worthy of study than performance; but

performance is also important and worthy of serious study, and it has its own significant regularities and principles (Hymes, 1974; Labov, 1972; Sternberg, 1988). Indeed, it is legitimate to focus on competence, performance, and the relation between the two. For example, if you are a teacher attempting to measure what your students have learned, you will focus on competence. If you are hiring an airplane pilot and want to know how well an applicant flies planes under pressure, you will focus on performance. If you are interested in motivation and why some people perform below potential, you will focus on the competence/performance discrepancy.

A second concern is that we have no fixed way of deciding what to label *competence* or *performance* (Newmeyer, 1983). Instead, what counts as competence and what counts as performance is relative to the perspective of the theorist. To a grammarian, rules of turn taking are performance; to one who studies pragmatics, they are part of competence. Even memory, usually assumed to be a performance factor, can be thought of as a competence that is influenced by aspects of performance such as mnemonic strategies (Smith, Sera, and Gattuso, 1988).

It is also important to keep in mind that the distinction between competence and performance is an idealization, and cannot be neatly assigned to distinct, independent processes (Blumstein, 1982). Some scientists suggest that the two may interact, so that changes in performance may affect competence (e.g., your competence in chess may suffer if you no longer practice the game). Competence can grow, change, or diminish over time, and as such is affected by opportunities to perform and practice.

EXAMPLES OF THE COMPETENCE/ PERFORMANCE DISTINCTION

For those outside linguistics, it has been useful to construe the distinction broadly, so that it includes competence beyond that of grammar and performance beyond that of speech. Given this more encompassing notion, competence/performance discrepancies are found whenever someone possesses knowledge. However, they are particularly striking when the knowledge is complex or rule-governed (such as language or

mathematics), or when the person whose knowledge is being assessed is operating under a different set of assumptions than the person testing the skill (for example, when adults study children, or anthropologists in the United States study people in another culture).

Consider first the skills of young children. Thirty years ago, parents were commonly told that their newborn infants could barely see, scarcely hear, and had no means of making sense of the chaotic jumble of sensations they experienced. Babies' behavior in daily activities and in psychological experiments supported this very limited portrayal. Yet it now seems that many of these difficulties resulted from performance limitations. The underlying competence is more advanced than would at first appear. In the last twenty years there has been a wealth of research revealing surprisingly sophisticated abilities in young children. We now know that infants, even newborns, can discriminate subtle speech sounds, recognize familiar odors, calculate the relative size of objects, and show strong preferences for particular visual patterns (such as faces).

Perhaps the most well-documented example of a competence/performance distinction in infancy concerns the development of object constancy: the understanding that objects continue to exist even when out of view. Jean Piaget (1954) observed that infants, if shown an attractive toy that is then placed behind a barrier in full view, either search for the toy in the wrong location or fail to look for it at all. Older children find the hidden toy with no difficulty. Piaget and others interpreted this result to mean that infants believe that objects no longer exist when out of sight. However, more probing techniques reveal that babies do keep hidden objects in mind (Harris, 1983). By studying infants' gazes, their patterns of reaching, or their reactions to seemingly impossible event sequences (such as two objects occupying the same place at the same time), researchers have found that infants in the first half year of life expect that objects continue to exist when hidden behind a screen. Even when babies cannot successfully find hidden objects, more subtle measures suggest that they know where the objects are.

Competence/performance discrepancies abound even with older children, who are not so behaviorally lim-

ited as infants. Some theorists (again, most notably Piaget) argued that preschoolers have fundamental limitations in their intellectual functioning. For example, two- to four-year-olds were described as not forming true concepts, not being able to take the perspective of another person, and not realizing that quantities remain the same across superficial transformations (e.g., the number of pennies I have does not change just because I stack them up in a tower instead of spreading them out). As with studies of infants, recent research suggests that many apparent limitations of preschool children reflect performance difficulties. Simpler, more focused tasks reveal that children have at least partial abilities in all these domains (Gelman, 1978).

Competence/performance distinctions also arise in studies of communication in other cultures (Hymes, 1974). Cultures differ in their conventions for answering questions appropriately. For example, among some English-speaking Native Americans it is inappropriate to answer a question without careful consideration (Philips, 1970). In these classrooms, schoolchildren meet a teacher's questions with silence. Similarly, children who have not yet learned the conventions for hand raising to obtain a turn may not have an opportunity to reveal their knowledge in a classroom. Split-second differences in the timing of a student's raised hand will lead others (even seven-year-old schoolchildren) to interpret the student as incompetent (McDermott & Tylbor, 1983).

An area of serious practical concern that raises competence/performance issues is that of IQ. The IQ is simply a test score, not a direct window onto either underlying intelligence or real-world behavior. Thus, from one perspective it taps performance (behavior on the test) and is an incomplete assessment of competence (underlying intelligence). From another perspective, it taps competence (knowledge of a particular limited sort, such as vocabulary and ability to solve analogy problems), not performance in school or success in the real world. From either perspective, the test is limited. By itself this is not a problem, yet difficulties arise when people confuse the test with the skills and abilities it is meant to measure. See Gardner (1983) and Sternberg (1988) for extended discussion of these issues.

SOURCES OF THE COMPETENCE/ PERFORMANCE DISCREPANCY

It is often difficult to determine why performance errors arise, and the issue must be addressed on a case-by-case basis. For example, ordinarily people make many errors in logical reasoning. Do these reflect difficulties in competence or in performance alone? According to one school of thought, they are strictly performance errors: We know the rules of logic but do not always apply them accurately. One demonstration in support of this point is that children's errors on a transitive-inferences task (*A* is greater than *B*, *B* is greater than *C*; is *A* greater than *C*?) drop significantly when the children are trained to remember the premises (*A* is greater than *B*, *B* is greater than *C*) (Bryant & Trabasso, 1971). Here, memory limitations blocked knowledge of an intact logical principle. However, other researchers argue that we do not possess abstract logical rules at all. Based on finding consistent, systematic patterns of performance errors and performance successes, they suggest that humans ordinarily use an alternative logic, following rules of reasoning that are different from the formal rules posited by logicians (Nisbett, Fong, Lehman, & Cheng, 1987).

How deep are performance limitations? The answer to this question partly depends on their sources. Performance errors are rooted in an extremely diverse set of factors, including cognitive constraints on how we process information (e.g., memory lapses, attentional deficits), personality traits (e.g., impulsivity, shyness), physical or emotional state (fatigue, anxiety), momentary distractions, cultural dictates, motivation to perform, and more. Some of these factors are superficial or fleeting, whereas others are virtually unchangeable. Most of these factors by themselves do not affect performance; instead they interact with the task to yield errors. For example, a shy person could score poorly on an oral exam but brilliantly on a written exam, whereas an illiterate person could show the reverse pattern. Thus, even for factors that appear to be more-or-less fixed (e.g., memory constraints, personality), tasks can be devised to bypass their effects and, consequently, to assess competence more adequately.

For the most part, sophisticated performance implies a high degree of competence. For example,

someone who can produce grammatical sentences in English probably has command of English grammar. However, performance can be misleading, at times implying an underlying competence that does not exist. For example, one person may prop up the performance of another (e.g., a parent deftly anticipating a child's reactions, so that together they carry out a conversation). Another example is the use of short-cut heuristics that closely approximate more in-depth knowledge. Alternatively, one may have encapsulated knowledge that is sophisticated in only a limited domain or context (e.g., the ability of bees to conduct complex navigations with respect to food finding but no other task) (von Frisch, 1950).

CONCLUSIONS

In summary, actual behavior and underlying knowledge are distinctly different things. In the words of Roger Brown (1973, p. 88), "samples of performance constitute only an imperfect set of clues to competence." Although only behavior is directly observable, with enough care and ingenuity it is possible to discover much about the structure of implicit knowledge, and to discover the factors that impede or assist performance. The distinction is important to keep in mind, whether one's focus is on competence, performance, or the interrelations between competence and performance.

BIBLIOGRAPHY

- BLUMSTEIN, S. E. (1982). Language dissolution in aphasia: Evidence for linguistic theory. In L. K. Opler & L. Menn (Eds.), *Exceptional language and linguistics* (pp. 203–215). New York: Academic Press.
- BROWN, R. (1973). *A first language*. Cambridge, MA: Harvard University Press.
- BRYANT, P. E., & TRABASSO, T. R. (1971). Transitive inferences and memory in young children. *Nature*, 232, 456–458.
- CHOMSKY, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- CHOMSKY, N. (1969). *Topics in the theory of generative grammar*. The Hague: Mouton.
- CHOMSKY, N. (1980). *Rules and representations*. New York: Columbia University Press.
- GARDNER, H. (1983). *Frames of mind*. New York: Basic Books.
- GELMAN, R. (1978). Cognitive development. *Annual Review of Psychology*, 29, 297–332.
- HARRIS, P. (1983). Infant cognition. In M. M. Haith & J. J. Campos (Eds.), *Infancy and developmental psychobiology*. Vol. 2: *Handbook of child psychology* (pp. 689–782). New York: Wiley.
- HYMES, D. H. (1974). Competence and performance in linguistic theory. In R. Huxley & E. Ingram (Eds.), *Language acquisition, models and methods* (pp. 3–23). New York: Academic Press.
- LABOV, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- MCDERMOTT, R., & TYLBOR, H. (1983). On the necessity of collusion in conversation. *Text*, 3, 277–297.
- NEWMAYER, F. J. (1983). *Grammatical theory: Its limits and its possibilities*. Chicago: University of Chicago Press.
- NISBETT, R. E., FONG, G. T., LEHMAN, D. R., & CHENG, P. W. (1987). Teaching reasoning. *Science*, 238, 625–631.
- PHILIPS, S. U. (1970). Acquisition of rules for appropriate speech usage. In J. E. Alatis (Ed.), *Georgetown University Round Table on Languages and Linguistics 1970* (pp. 77–101). Washington, DC: Georgetown University Press.
- PIAGET, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- SMITH, L. B., SERA, M., & GATTUSO, B. (1988). The development of thinking. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought* (pp. 366–391). Cambridge: Cambridge University Press.
- STERNBERG, R. J. (1988). Intelligence. In R. J. Sternberg & E. E. Smith (Eds.), *The psychology of human thought* (pp. 267–308). Cambridge: Cambridge University Press.
- VON FRISCH, K. (1950). *Bees: Their vision, chemical senses, and language*. Ithaca: Cornell University Press.

SUSAN A. GELMAN

COMPLEXITY, COGNITIVE Intelligence is a concept that deals with intellectual power; cognitive complexity is a concept (actually, a group of related concepts) that deals with the ways people apply the intellectual power that they possess. Most theorists consider cognitive complexity to be a personality factor that varies from one person to the next but is stable within at least adult individuals. It affects one's approach to a wide range of situations and problems and constitutes a bias or preference in dealing with information.

COGNITIVE STRUCTURE

Complexity theories describe cognitive structure. This may be thought of metaphorically as the framework of a building, which influences and to some extent determines the arrangement and function of different rooms. The actual furnishings, decoration, and utilization of the rooms, however, can then vary widely according to the resources and needs of individual users. Cognitive structures serve the intellectual process regardless of content. They guide us in selecting information to be processed from the total available, organizing the selected information in some characteristic fashion, moderating and controlling the impact of emotion and motivation upon information processing, and adapting to situational requirements posed by particular tasks (Bieri, 1971). Their function is to help us cope with the tremendous amount of stimulation and information (some presented in the environment and some generated internally through memories, emotions, and the like) that bombard us most of the time. They guide us as to what information to ignore and what to process, and how to interpret the latter. This filtering function makes it easier to act upon the information and to identify that part of the information that requires the time-consuming and effortful application of "pure"—that is, relatively unbiased—intellectual processes.

The cognitive complexity approach is related, both theoretically and empirically, to other cognitive style constructs. All guide the functioning of intelligence, reasoning, logic, problem solving, and planning. Several share the concentration on structure rather than content. Complexity theorists have made a particular point of this emphasis. They have devoted themselves primarily to the study of two components of structure: differentiation and integration.

Differentiation and Integration. Differentiation refers to the perception that there is more than one relevant dimension or factor within a stimulus (person, event, concept, plan, problem, and so on) or to the recognition that there is more than one legitimate way of looking at or interpreting that stimulus. An undifferentiated cognitive structure may apply one metric (e.g., to my advantage/not to my advantage) to all events or plans of action. A differentiated one may consider such additional factors as long- versus short-

term outcomes, moral and legal acceptability, probability of success, and impact upon other people.

Integration is the relating of such differentiated cognitive units to each other. Integration can occur through the combination of several dimensions or perspectives, by the conception of trade-offs or compromises between them, or by thinking of them as components of a larger, superordinate cognitive schema or system (see Figure 1). Obviously, differentiation is a necessary precondition for integration, though not sufficient in itself.

Differentiation is a basic component in all measures and theories of cognitive complexity. Some theorists have elaborated the construct. For example, Scott, Osgood, and Peterson (1979) distinguished among the number of attributes an object is perceived to have (object complexity), the number of categories within an attribute (attribute precision), and the number and distinctness of attributes available to the person (dimensionality). Others focus on a particular kind of differentiation: P. W. Linville (e.g., 1987) and her colleagues have studied the attitudinal and health implications of "self-complexity," the degree to which one's cognitive representation of the self is differentiated.

Both differentiation and integration are relevant to complexity; each is unique, but they are related. Consideration of both seems necessary for a comprehensive analysis of complexity as a cognitive style. Nevertheless, integration is incorporated explicitly by only a few of the complexity models. The systems that deal most satisfactorily with integration are conceptual complexity theory (Schroder, Driver, & Streufert, 1967) and its offshoots. Among these is the integrative complexity approach, which studies complexity as a state variable expressed in a particular situation rather than as a cross-situationally stable trait (Suedfeld, Tetlock, & Streufert, 1992).

MEASURING COGNITIVE COMPLEXITY

As with any theoretical construct involving important individual differences, the measurement of cognitive complexity has attracted considerable attention. Because there have been several versions of complexity theory, there have also been a number of measurement procedures proposed and tried.

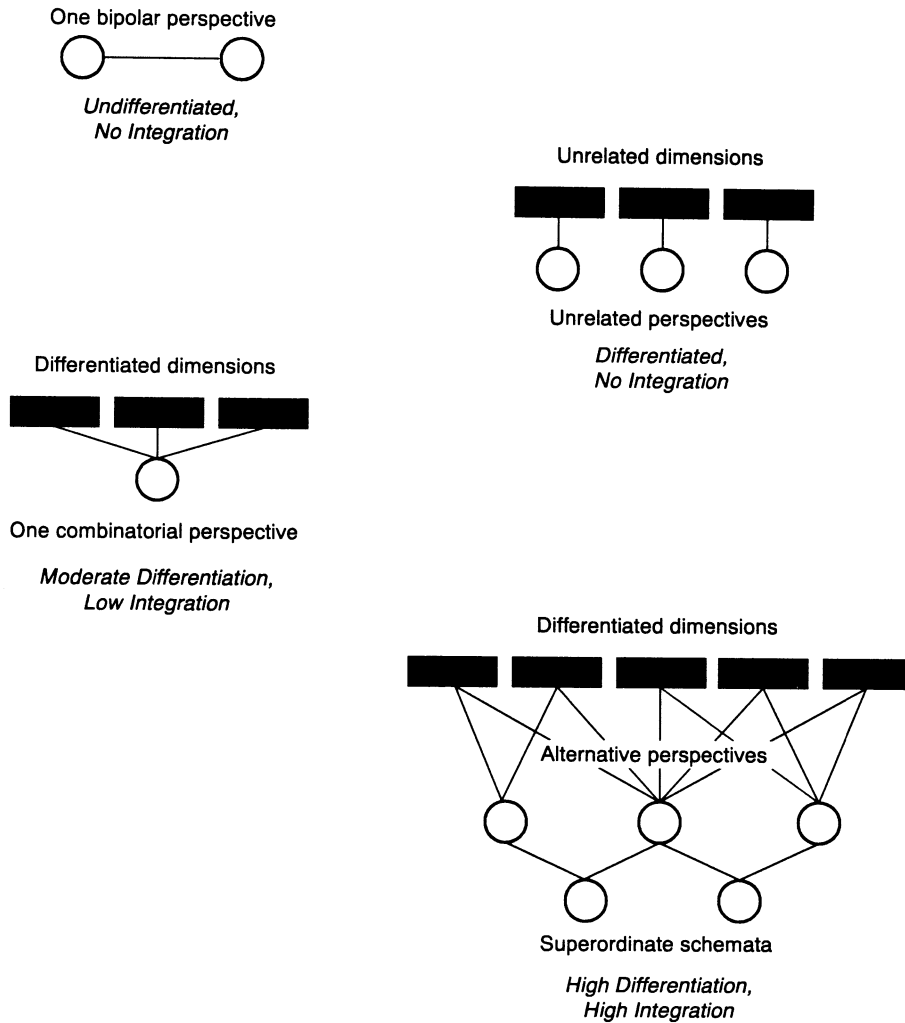


Figure 1
Differentiation and integration

The earliest of these, still used by researchers, is George Kelly's Role Construct Repertory Test (1955; more commonly known as the Rep Test). The procedure requires the respondent to write down the names of acquaintances who fit each one of several roles presented in the test. The subject is then asked to think of important ways in which any two of these people resemble each other but not a third. The basis of similarity is called the construct; the opposite of the construct, which differentiates between the first two individuals and the third, is the contrast. The subject then goes through all of the other acquaintances named and indicates whether each belongs in the construct or the contrast category (i.e., resembles or does

not resemble the two original people). Then a matrix is made up to compare the pattern of how people are rated on constructs. Highly similar ratings (i.e., many people rated the same way) across constructs show a simple cognitive structure. This is a measure of differentiation. Later modifications, such as factor analyzing the responses or having the researcher provide preselected constructs on which to rate stimuli, have been based on the same definition of cognitive complexity: the more different constructs of factors can be identified, or the more they differ from each other, the higher the level of complexity.

The major alternative to this kind of assessment measures both differentiation and integration. It usu-

TABLE 1
Integrative Complexity

Score	Characteristic
1	Undifferentiated, no integration
2	Transitional
3	Differentiated, no integration
4	Transitional
5	High differentiation, low integration
6	Transitional
7	High differentiation, high integration

ally involves the respondent's writing one or more sentences, paragraphs, essays, or passages on themes chosen by the tester (e.g., the Paragraph Completion Test, or PCT; Schroder, Driver, & Streufert, 1967). These are then analyzed by trained scorers for evidence of the two components of complexity and scored on a 1–7 scale (Table 1).

This method requires careful training and considerable work for scorers. To eliminate these drawbacks, attempts have been made to replace it by multidimensional scaling and multiple-choice formats. No valid and reliable substitute has been developed as of the early 1990s. On the other hand, the method itself has been improved and extended. The original PCT was a speed test, using six paragraph topics with 1 ½ to 2 minutes per paragraph. Newer formats include a power test allowing 8 minutes for each of two topics, and a scoring system that can be applied to documents, novels, letters, videotapes, interviews—in fact, to almost any connected verbal material (see Suedfeld, Tetlock, & Streufert, 1992).

COMPLEXITY AND INTELLIGENCE: STATISTICAL ASSOCIATIONS

The association between cognitive complexity and intelligence (as measured by standard IQ tests) is of theoretical interest. Is complexity in fact a facet of intelligence? Or, can higher levels of complexity be reached only by people who attain at least some level of verbal ability (and if so, what is that level)? Conversely, it can be argued that some intelligence test items may be measuring complexity, at least to some extent. In view of the reliance of complexity measures on verbal behavior, these seem to be reasonable ques-

tions with important implications for the complexity theories.

Correlation studies have shown that although some measures of differentiation and intelligence are correlated, one is not merely a version of the other. Kelly's Rep Test and its variants have typically shown low correlations (around .20) with either IQ scores or academic grades, although there have been a few exceptions with correlations up to the .40s. One of the highest correlations was obtained with a group of low-IQ subjects (see Goldstein & Blackman, 1978). Scott, Osgood, and Peterson (1979), using a much more elaborate but conceptually similar listing and sorting task, found correlations with Scholastic Aptitude Test scores in the .2 to high .3 range, depending upon the domain being differentiated.

With the Paragraph Completion Test, which measures differentiation and integration, the association depends upon the subject group. Adults show mixed results, with correlations ranging from the .20s to the .40s; again, the correlation is higher among less intelligent groups. With subjects of high school age or younger, the correlations are consistently positive and significant. Suedfeld and Coren (1992) found that the power PCT (two paragraphs, no time pressure) correlated reliably with measures of divergent thinking but not with fluid or crystallized intelligence.

It appears that there is an association between cognitive complexity and at least verbal intelligence if one looks at a sufficiently broad range of IQs. This relationship is particularly strong among groups whose verbal facility may not be fully developed, such as children and low scorers on the intelligence measure. The underlying reasons may be both that people must reach a threshold of verbal skill to express highly differentiated or integrated concepts and also that some indices of IQ in effect require differentiation or integration.

COMPLEXITY AND INTELLIGENCE: PROBLEM SOLVING

The substantive, as opposed to the statistical, association between intelligence and cognitive complexity is seen when complexity is considered as a factor in how people solve problems. This is also the crucial test of what it means to be intelligent.

Cognitive complexity affects the processing of all information and therefore all forms of thinking and decision making. Researchers have documented its role in a wide variety of contexts from the solving of experimental puzzles, through adaptation to a new culture and the interaction between teachers and pupils, to the strategies of governmental leaders facing international crises.

Complex people are, predictably, more complex in solving problems. They collect and consider more of the information available in a situation and put together solutions in which each step is more connected to the others at any one point as well as over time. Complex managers see more links between their own decisions and ongoing events and use these links to guide further information search; they originate more options in problem solving; and they pay more attention to the reactions of the people with whom they are dealing (Streufert & Swezey, 1986). Complex people are also better at communicating in such situations as marital problems, cross-cultural cooperation, and teacher-pupil and counselor-client interactions (for references, see, e.g., Suedfeld, Tetlock, & Streufert, 1992). They are also less likely to develop depression, physical illness, and low self-esteem under stress or after failure.

Complexity is not necessarily beneficial: high levels of information search and processing can be self-defeating. Complex subjects have greater difficulty in making choices and decisions, may be misled by paying too much attention to incorrect or irrelevant information, and may take too much time in reaching a conclusion (Goldstein & Blackman, 1978). They may seem vacillating and indecisive to others (Tetlock, 1991). They are also more aroused physiologically by some challenging situations, more anxious, and more likely to have gastrointestinal dysfunctions and heart attacks (Streufert & Swezey, 1986).

In the case of complex behavior (as opposed to complex personality), different levels of differentiation and integration seem optimal in different circumstances. Sometimes the situation calls for clear, straightforward actions; sometimes, for compromise, negotiation, and empathy. Individuals who can sense which is appropriate and respond accordingly do better than those who stay rigidly at one level, whether

high or low. Flexible complexity predicts, for example, long-term career success in political leadership and international diplomacy (Suedfeld, Tetlock, & Streufert, 1992).

CONCLUSION

Cognitive complexity is itself a complex idea. The phrase can refer to a personality trait—that is, to a consistent way in which the individual approaches intellectual problems and decisions. Alternatively, it can mean the way in which that approach is made in a specific situation. It shares some theoretical and statistical ground with the traditional concept of intelligence, but its closest association with that concept is in how both relate to the person's ability to cope with information and with life.

BIBLIOGRAPHY

- BIERI, J. (1971). Cognitive structures in personality. In H. M. Schroder & P. Suedfeld (Eds.), *Personality theory and information processing* (pp. 178–199). New York: Ronald.
- GOLDSTEIN, K. M., & BLACKMAN, S. (1978). *Cognitive style: Five approaches and relevant research*. New York: Wiley.
- KELLY, G. A. (1955). *A theory of personal constructs*. New York: W. W. Norton.
- LINVILLE, P. W. (1987). Self-complexity as a cognitive buffer against stress-related illness and depression. *Journal of Personality and Social Psychology*, 52, 663–676.
- SCHRODER, H. M., DRIVER, M. J., & STREUFERT, S. (1967). *Human information processing*. New York: Henry Holt.
- SCOTT, W. A., OSGOOD, D. W., & PETERSON, C. (1979). *Cognitive structure: Theory and measurement of individual differences*. Washington, DC: Winston.
- STREUFERT, S., & SWEZEY, R. W. (1986). *Complexity, managers, and organizations*. New York: Academic Press.
- SUEDFELD, P., & COREN, S. (1992). Cognitive correlates of conceptual complexity. *Personality and Individual Differences*, 13, 1193–1199.
- SUEDFELD, P., TETLOCK, P. E. & STREUFERT, S. (1992). Conceptual/integrative complexity. In C. P. Smith (Ed.), *Motivation and personality: Handbook of thematic content analysis*. Cambridge: Cambridge University Press.
- TETLOCK, P. E. (1991, August). An integratively complex look at integrative complexity. In P. Suedfeld (Chair),

New directions in research on integrative complexity. Symposium conducted at the annual meeting of the American Psychological Association, San Francisco.

PETER SUEDFELD

COMPREHENSION See WAIS-R SUBSCALES.

CONNECTIONISM Much of the research and theorizing within cognitive science is based on the assumption that the mind, like a computer, is a general information-processing device. This assumption implies that both the mind and the computer perceive information from an input device (e.g., eyes, ears, touch) and, through a series of processing steps, evaluate and transform the information for understanding and/or output. The adoption of this perspective has led to the investigation of information-processing characteristics of such systems as computers and minds (Newell & Simon, 1972).

Two basic premises underlie these types of information-processing systems: The systems operate on physical symbols (information, data), and these operations are defined as the manipulation of symbols by rules in the form of programs (Newell & Simon, 1972). The physical-symbol premise asserts that each concept is uniquely and independently represented, that is, there is a specific data item in storage that stands for a given concept. The program (rules) takes the data (symbols) and manipulates and transforms them to produce some new output. The theoretical consequence of adopting these premises has been the attempt to specify the "program," the "software," or the rules that people at different levels of skill or ability use to manipulate data and how rules develop to account for performance.

Connectionism offers a different perspective on how to characterize the mind. Connectionist models are a class of models that share a set of underlying assumptions (Feldman & Ballard, 1982; Rumelhart, Hinton, & McClelland, 1986). Connectionist models also posit that the mind is an information-processing device. However, these models suggest that the mind

processes information not like a computer but like neurons in the brain. The brain is composed of neurons, which process information through the transmission of neurotransmitters from the axon terminal of one neuron, across the synapse, to the dendrite of another neuron. These neurotransmitters act to change the polarization, or electrical potential, of the receptor neuron.

The assumption that the mind is like the brain leads psychologists to make very different assumptions regarding how the mind works. As will be discussed, the brain orientation differs from the computer orientation in assuming that information processing is not carried out via the process of symbol manipulation and that knowledge is best conceptualized in terms of distributed representations, not physical symbols. Thus, connectionist models assert that the two main assumptions underlying the simile that "the mind is like a computer" are inaccurate.

Connectionist models posit that information is processed by simple processing units that are interconnected by associative links. These units behave like neurons. The premise is that each of these processing units has associated with it some level of activation. The level of activation might be high enough so that the unit can be considered "on" or low enough so that it can be considered "off." A second premise is that each of these units is interconnected with other units by associative links. When a unit is active, it sends activation across the associative links, and this affects the activation level of its neighbors. Within connectionist models, two different types of associative links exist. The first is "excitatory links," which increase the activation level of the neighboring units, and the second is "inhibitory links," which decrease the level of activation of the units. This process of sending activation across associative links has been labeled "spreading activation." Each unit has the capability of receiving as well as sending activation. A given unit's activation level will ultimately be determined by evaluating all the messages that it receives from the environment and from its neighbors (both excitatory and inhibitory). An important component of connectionist models is that the spread of activation across associative links occurs in parallel; that is, many different units can send and receive activation at the same time.

One way to think about these models is to equate each processing unit to a demon. Each demon has the capacity to yell (level of activation). It can be quiet (resting), yell very loudly, or call out at some level in between. In addition, once a demon starts yelling, it sends messages to demons associated with it (across associative links), either telling them to yell (excitatory messages) or telling them to keep quiet (inhibitory messages). Many different demons can be yelling and sending messages at the same time. Each demon who receives these messages evaluates all the messages coming in (each demon can receive messages from many different demons) and then determines how much it should yell. Given a certain input (environment, state of problem), one can envision a chorus of devils yelling and many keeping quiet. Ultimately, each demon will determine how much it should yell, and a certain level of yelling across demons will be obtained.

It is important to note that unlike the computer perspective, which asserts the use of a physical symbol, the connectionist view asserts that no one demon (unit) means anything alone. Meaning is derived from the pattern of yelling between demons. Any one demon may be part of many different groupings. This is what is meant by a “distributed representation.”

The next issue to be addressed is how information processing is carried out in such systems. Most connectionist models involve at least three different levels of units (demons). The first level of units can be characterized as representing the stimulus or environment. The second level is characterized as an inference level, which represents the person’s understanding of the environment. The third level is characterized as an output level and represents hypotheses regarding different ways of responding. Thus, there is a pattern of activation across units that represents the input, a pattern of activation across units that represents understanding, and a pattern of activation across units that represents the response.

In most situations, the manner in which these levels interact is clearly ordered. Activation starts at the level that represents the input, which then activates the level that represents understanding, which in turn activates the level that offers responses. This description greatly simplifies how these levels interact. First, it is not necessary for one level to settle down to a pattern of activation prior to activating another level. Second,

the first level can have direct links to both the second and third levels, so that activation can flow from the first to the third level directly, as well as from the second to the third level. Third, activation can also flow downward through the levels, so that the pattern of activation on the higher levels can influence the pattern on lower levels. It is important to realize that these levels are highly interactive.

Finally, while the output level can represent some desired behavior, it does not have to. The output level of any network can be thought of as a thought or an idea that a person thinks of while working on something. The output level, then, can be an internal state. This output level then serves as an input level for another network. Thus, information processing is conceptualized as one pattern of activation across units affecting another pattern. Processing is not carried out by rules or programs that manipulate data; it is carried out by spreading activation that changes activation levels of related units.

The importance of connectionist models is that they provide a different perspective for understanding cognition and intelligent performance. From the perspective of the simile that “the mind is like a computer,” competent performance is seen in terms of the acquisition and development of processes that permit more efficient transformations of data. From the connectionist perspective, competent performance is understood in terms of the development of networks, which do not explicitly represent rules. The understanding of intelligence and skilled performance involves understanding the constraints on how these networks develop, not how rules are acquired.

BIBLIOGRAPHY

- FELDMAN, J. A., & BALLARD, D. H. (1982). Connectionist models and their properties. *Cognitive Science*, 2, 205–254.
- NEWELL, A., & SIMON, H. A. (1972). *Human problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- RUMELHART, D. E.; HINTON, G. E.; & MCCLELLAND, J. L. (1986). A general framework for parallel distributed processing. In D. E. Rumelhart & J. L. McClelland (Eds.), *Parallel distributed processing* (Vol. 1, pp. 45–76). Cambridge, MA: MIT Press.

MITCHELL RABINOWITZ

CONSERVATION See CHILDREN'S CONSERVATION CONCEPTS.

CONTEXTUALIST THEORIES OF INTELLIGENCE All theories of human intelligence are necessarily contextualist. They differ, sometimes strikingly, with respect to which aspects of context are viewed as most central for understanding human intelligence. Theories of intelligence that place the greatest emphasis on the HERITABILITY of individual differences in intelligence assert the importance of evolutionary history—and thus the contextual factors that operate through natural selection—in shaping the distribution of genetic factors that influence observed (phenotypic) diversity (Eysenck, 1988; Galton, 1892/1962). In contrast, theories of intelligence that focus on the diversity arising from an individual's developmental history emphasize the contextual contingencies operating in the person's physical and social environment (Hunt, 1961).

Although the need to integrate contrasting theoretical approaches has long been recognized (Anastasi, 1958; Dewey, 1922), recent theoretical and research efforts have sought new ways to bridge the gap in an attempt to understand human intelligence by incorporating these disparate contextual effects (Bronfenbrenner & Crouter, 1983; Ceci, 1990; Cole & Scribner, 1974; Gardner, 1983; Sternberg, 1990). Several clarifications may help to avoid confusions that have hindered previous integrative attempts.

The traditional dichotomy between nature and nurture—that is, inheritance or genetic accounts versus environmental or experiential accounts—has often obstructed understanding. The approach of many investigators to the study of mind was limited by the methodologies available and accepted at the time. These tended to be main effect models (Green, 1992). From this perspective, the essence of intelligence can best be found by looking for the central effects while making all else equal. In this formulation, the interesting story is what remains after the effects of individual histories have been removed through various statistical forms of equating.

For the individual organism, of course, these influences are not dichotomous but rather are fully integrated during development. Particularly relevant here

is the impact of experiential history on the sculpting of neural, immune, and hormonal patterns. Researchers have devoted much effort to the division of effects between these competing factors, a goal that becomes more complex and nuanced as the precision of these estimates increases (Plomin & Thompson, 1988).

Historically, relatively less effort has gone toward the construction of robust developmental accounts that can explain how these two substantive influences interact to yield the observed diversity in human intellectual performance. As a result, researchers have shifted their focus beyond the strict apportionment of isolated effects toward the more complicated task of describing the dynamic interaction of multiple influences over the course of human development.

Some past confusion derives from the failure to distinguish between contextual factors associated with intellectual development in general and those associated with diversity in intellectual accomplishment. A useful distinction separates capacities and capabilities, a distinction masked by the omnibus term "mental abilities." Literacy offers a helpful example. The vast majority of humans obviously have the capacity to become literate, given the appropriate experiential contingencies. Previously illiterate populations demonstrate high proportions of literacy with the advent of schooling, rapidly becoming capable of reading.

Because of theoretical assumptions prevailing in the early history of empirical research on human intelligence (e.g., Terman, 1916), these two constructs—capacity and capability—were conflated, in the belief that attained capabilities were quite reliable estimates of fundamental intellectual capacity (Keating, 1990a). More recent efforts to disentangle these notions, in order to achieve purer, more reliable estimates of fundamental capacity independent of developmental influences—such as information-processing capacity or efficiency—have generated mixed results. Again, the processes of developmental integration may make it difficult to disentangle these issues. This situation reinforces the need to examine in much greater detail the nature of human diversity as a developmental phenomenon.

Such an examination logically begins by focusing on those features of intelligence that arise out of our evolutionary history. As Howard Gardner (1983) has argued, this understanding has important implications

for contemporary theories of intelligence with respect to issues like modularity, domain specificity or generality, and hypothesized pathways for the expression of genetic variation. Understanding the ecological contexts that influenced natural selection is thus important not only in its own right, but also in terms of the light it may shed on the contemporary development of human intelligence.

Two areas of evolutionary investigation are particularly relevant—the characteristics that are shared with our primate cousins and those that appear to be distinctly human. Greater attention has typically concentrated on the latter, but recent evidence from comparative ethology, together with the degree of genetic overlap among human and nonhuman primates, suggests that such studies may be a useful source for understanding the characteristics of our own species (Suomi, 1991). In particular, the roles of group cohesion, social hierarchies, and interpersonal attachments and alliances are probably central for primate development in general (Suomi, 1991; Tomasello, Kruger, & Ratner, 1993). To the extent that social interaction is a key component of human intellectual development, this overlap can provide important information on some of the biological constraints that affect it (Dunbar, 1992).

A variety of explanations account for hominid speciation (Stringer & Gamble, 1993). Given the difficulty of obtaining hard evidence on this question, a priority among various explanations does not exist. That the use of language is a defining species characteristic of humans is not controversial. In contrast, the origins of language, the magnitude of discontinuity in primate evolution it represents, and its range (isolated or broadly integrated) as a biological accomplishment all generate competing explanations. Similar controversy surrounds the distinctiveness of human tool use, compared with other primates. L. S. VYGOTSKY (1978) proposed that neither language nor tool use alone was the origin of human intelligence, but rather the combination of these skills (Wertsch, 1985).

Stephen Jay Gould (1981, 1982) has suggested a more general shift as the basis of specifically human intelligence. What is different about humans, and in many ways the most significant characteristic of the species, is the long period of developmental plasticity, during which mental structures emerge in close at-

tunement with the local environment. This feature, termed *neoteny*, characterizes an animal capable of learning broader and more flexible than that of other species. From this perspective, the relationship between experience and intelligence is neither direct nor one-way. The mental structures that evolve as a result of experiential contingencies, even from a very early age, are immediately influential in the type of available information that the organism attends to and in the interpretation placed upon that information.

Two key consequences arise from the long period of plasticity in human learning. First, diversity within the population is to be expected. Second, accumulated cultural knowledge becomes the cognitively socializing habitat of the present. Thus, cultural learning in the sense of intellectual collaboration is a uniquely human form of learning and a likely source of cultural evolution (Tomasello, Kruger, & Ratner, 1993).

Whatever the final resolution of these claims about the processes of natural selection in hominid evolution, the contexts that shape human intelligence have a clear, dramatic shift with the onset of cultural evolution. Two examples illustrate the breadth of these changes. Literacy has become an obvious central factor defining the success of modern societies. The onset of literacy marks a shift, however, not only in the way in which communication can occur but also in the kinds of collaborative learning that are possible and the types of mental structures and processes that then become available (Olson, 1977; Scribner & Cole, 1981).

Similarly, the shift from concrete to abstract representation and categorization, long seen as a key indicator of individual intellectual development, may have a close link to the ways of thinking generated by formal schooling (Luria, 1976; Sharp, Cole, & Lave, 1979).

How strongly social and cultural contexts affect fundamental forms or structures of human intelligence, compared with a more peripheral shift in content knowledge and skills, remains unanswered (Sternberg, 1990). At one extreme are models of maturational unfolding of intellectual structures, for which the context either affords (or fails to afford) adequate sustenance. At the other extreme are models that stress more or less direct inculcation of environmental features into mental activity. In reality, productive theories of human intelligence have not occupied either

extreme position. All serious theories of human intellectual development afford a substantial role to the influence of contexts. They differ, however, in the degree to which contextual variation is incorporated into the fundamental structures of intelligence.

Jean PIAGET (1950) maintained a key role for the influence of context on the expression of intellectual functions, ascribing cognitive development to ongoing, reciprocal interactions between the individual and the environment, both physical and social. These interactions, however, presumably generate a common progress in logical stages, despite differences among individuals in their interactive context. Contemporary critiques of the Piagetian approach have focused on this claim to universality of the emergent logical structures. The major empirical difficulties for the Piagetian model are the apparent domain specificity of cognitive development and the lack of within-person consistency in logical achievements.

Neo-Piagetian models (Case, 1992; Fischer & Pipp, 1984) have sought to reconcile this conflict by combining two features: general or systemwide cognitive developmental constraints, which set overall stagelike limits, and local or domain-specific constraints that are far more sensitive to individual developmental history and context. The success of such dual system models remains uncertain, but the movement toward more contextualized versions of general intellectual development is clear.

Compared with the Piagetian approach, Vygotsky (1978) emphasized the role of the social context in shaping the nature of mental activity (Wertsch, 1985). As noted above, he focused on the unification of symbol and tool use as the origin of truly human intelligence, both for the species and the individual. This unification occurs through the internalization of external discourse. Intellectual development proceeds most efficiently when that discourse is scaffolded to provide experiences within the individual's zone of proximal development, which ranges from what persons can accomplish unaided to what they are able to do with external supports. Along with Piaget, Vygotsky emphasized qualitative discontinuities in the acquisition of mature intelligence, but he also believed that the timing and nature of these discontinuities are largely idiosyncratic, reflecting the diversity among individuals in the contexts for learning.

Historically, the greatest controversy over the role of context has arisen from the explanation of diversity among individuals. The Galtonian tradition holds that individual differences in intelligence are largely biological and that the influence of context in individual development arises peripherally, either through deprivation that suppresses the expression of intelligence or through the impact of different contexts on the contents (but not the essence) of intellectual activity (Eysenck, 1988). Environmental theorists have long disputed this claim, pointing to the impact of different rearing conditions on intelligence outcomes (e.g., Hunt, 1961). Given the stark contrast between these two views of the world, it is not surprising that the conflict between them has been heated in all spheres—scientific, political, and cultural. In order to move toward a more integrated and productive scientific perspective, key weaknesses of the bipolarity need to be addressed.

Much of the early debate focused on the nature of general intelligence. It was assumed that some intellectual essence was distributed differentially across persons or in some cases, across species (Galton, 1892/1962) and that this essence would emerge more or less intact, given roughly equivalent rearing conditions. Two kinds of evidence are supportive of this position. Analyses of test performance yielded a strong general factor, and behavior genetic findings supported a moderate to high estimate of heritability of intelligence. Well-known difficulties in the mathematical specificity and psychological interpretability of factor analytic evidence (Glymour et al., 1987; Keating & MacLean, 1987; Sternberg, 1990), together with more complex and differentiated behavior genetic evidence (Plomin & Thompson, 1988), have suggested that a more complex model is needed. This has reinvigorated researchers' interest in the role of development and context in the generation of intellectual diversity.

A consequence of this renewed interest is a return to the question of intelligence as a general or specific quantity. Gardner (1983) has advanced a MULTIPLE INTELLIGENCES THEORY, which takes factor-analytic findings of performance as only one of multiple criteria necessary in determining the issue. Other key criteria include the coherence of development within an hypothesized intelligence and a plausible evolutionary story.

Many studies of social intelligence and social cognition have failed to find any construct valid differentiation between academic-type intelligence measures and social intelligence measures. The parsimonious interpretation seems to be that these tasks correlate substantially with most other traditional cognitive and intellectual tasks, weakening the case for differential domains. But there is a major empirical difficulty with this parsimonious interpretation: Success on social-intelligence tasks appears to be almost wholly unrelated to social competence, if we define social competence as doing well in real-world social situations. Similar difficulties plague other hard-to-measure human competencies, such as emotional awareness (Oatley & Jenkins, 1992). Indeed, the relative difficulty of measuring the socioemotional domain in a way that captures its real diversity, rather than reducing it to easily measurable dimensions, has tended to exclude it from concerns about human intelligence. Such exclusion appears to derive from assumptions about what aspects of human intelligence are central. Which aspects are indeed central is an interesting and open question. A wide range of evidence needs to be brought to bear, including evolutionary, sociohistorical, and developmental findings.

Whether the dichotomy between domain-general versus domain-specific intellectual development is productive thus depends partly upon the criteria used to investigate and determine specificity. If they are sufficiently developmental and contextual, the theoretical tension may be valuable (Keating & Crane, 1990; Sternberg, 1989). For example, the habits of mind that develop from early socioemotional experiences may have quite general effects on subsequent attentional patterns, and thus on intelligence (Lewis, 1993). Aspects of more formal cognitive socialization, especially the context of schooling, are apparently more important for the acquisition of specific expertise. Ceci (1990, 1991) has amassed extensive evidence that shows the impact of specific contexts on a wide range of intellectual performance. Using a combination of case study and experimental evidence, M. Howe (1990) has argued for substantial contextual influences in the formation of exceptional abilities. In all cases, successful integration of the contexts of intellectual development into a coherent story of the diversity of intellectual outcomes must account more precisely for

the specifics of the cognitively socializing habitat in relation to the particular period of development in which they are presumed to operate.

The major challenge confronting the study of human intelligence is thus the necessity of combining the disparate levels of contextual influence into a cogent account of intellectual development. Several promising new avenues give reason for optimism. First, the available analytic and methodological models for studying cognitive development in relation to its contextual settings have become more sophisticated, making a wider range of relevant questions possible, including emerging models for the investigation and modeling of dynamic systems. Second, the shift from exclusive consideration of demographic indicators to estimate environmental effects toward the microlevel social contexts and interactions that are responsible for developmental transmission, focuses research attention where the contextual action really lies (Bronfenbrenner & Crouter, 1983; Keating & MacLean, 1988). Third, the emergent crisis within education regarding the system's ability to deal with developmental diversity also focuses attention in a highly practical way on the validity and utility of our prevailing models of intelligence, learning, and diversity (Keating, 1990a).

Whether the current efforts to integrate context meaningfully into theories of human intelligence will be more successful than previous attempts remains uncertain. Any comprehensive picture of human intelligence will, however, need to include a coherent and defensible story of the multiple contexts in which intellectual development and intellectual diversity occur.

(See also: INTERACTIONIST VIEWS ON INTELLIGENCE.)

BIBLIOGRAPHY

- ANASTASI, A. (1958). Heredity, environment, and the question "How?". *Psychological Review*, 65, 197-208.
- BRONFENBRENNER, U., & CROUTER, A. C. (1983). The evolution of environmental models in developmental research. In P. H. Mussen (Ed.), *Handbook of child psychology* (4th ed., Vol. 1, pp. 357-414). New York: Wiley.
- CASE, R. (1992). *The mind's staircase: Exploring the conceptual underpinnings of children's thought and knowledge*. Hillsdale, NJ: Erlbaum.

- CECI, S. J. (1990). *On intelligence . . . More or less: A bio-ecological treatise on intellectual development*. Englewood Cliffs, NJ: Prentice Hall.
- CECI, S. J. (1991). How much does schooling influence cognitive and intellectual development? A reassessment of the evidence. *Developmental Psychology*, 27, 703–722.
- COLE, M., & SCRIBNER, S. (1974). *Culture and thought: A psychological introduction*. New York: Wiley.
- CYNADER, M., SHAW, C., PRUSKY, G., & VAN HUIZEN, F. (1990). Neural mechanisms underlying modifiability of response properties in developing cat visual cortex. In B. Cohen & I. Bodis-Wolliner (Eds.), *Vision and the brain: The organization of the central visual system*. New York: Raven.
- DEWEY, J. (1992). *Human nature and conduct*. New York: Henry Holt.
- DUNBAR, R. (1992). Why gossip is good for you. *New Scientist*, 136, 28–31.
- EYSENCK, H. J. (1988). The concept of “intelligence”: Useful or useless? *Intelligence*, 12, 1–16.
- FISCHER, K. W. (1980). A theory of cognitive development: The control and construction hierarchical skills. *Psychological Review*, 87, 417–431.
- FISCHER, K. W., & PIPP, S. L. (1984). Processes of cognitive development: Optimal level and skill acquisition. In R. J. Sternberg (Ed.), *Mechanisms of cognitive development* (pp. 45–75). New York: W. H. Freeman.
- GALTON, F. (1962). *Hereditary genius: An inquiry into its laws and consequences*. Cleveland, OH: World. (Original work published 1892.)
- GARDNER, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- GELMAN, R., & BAILLARGEON, R. (1983). A review of some Piagetian concepts. In J. H. Flavell & E. M. Markman (Eds.), *Handbook of child psychology* (Vol. 3, pp. 167–230). New York: Wiley.
- GLYMOUR, C., SCHEINES, R., SPIRITES, P., & KELLY, K. (1987). *Discovering causal structure*. Orlando, FL: Academic Press.
- GOULD, S. J. (1981). *The mismeasure of man*. New York: W. W. Norton.
- GOULD, S. J. (1982). Changes in developmental timing as a mechanism of macroevolution. In J. Bonner (Ed.), *Evolution and development*. Berlin: Springer-Verlag.
- GREEN, B. F. (1992). Exposé or smear? The Burt affair. *Psychological Science*, 3, 328–331.
- HOWE, M. J. A. (1990). *The origins of exceptional abilities*. Oxford: Basil Blackwell.
- HUNT, E. (1978). Mechanics of verbal ability. *Psychological Review*, 85, 109–130.
- HUNT, J. MCV. (1961). *Intelligence and experience*. New York: Ronald Press.
- KEATING, D. P. (1990a). Charting pathways to the development of expertise. *Educational Psychologist*, 25, 243–267.
- KEATING, D. P. (1990b). Structuralism, deconstruction, reconstruction: The limits of reasoning. In W. F. Overton (Ed.), *Reasoning, necessity, and logic: Developmental perspectives*. Hillsdale, NJ: Erlbaum.
- KEATING, D. P., & CRANE, L. L. (1990). Domain-general and domain-specific processes in proportional reasoning. *Merrill-Palmer Quarterly*, 36, 411–424.
- KEATING, D. P., LIST, J. A., & MERRIMAN, W. E. (1985). Cognitive processing and cognitive ability: A multivariate validity investigation. *Intelligence*, 9, 149–170.
- KEATING, D. P., & MACLEAN, D. J. (1987). Cognitive processing, cognitive ability, and development: A reconsideration. In P. A. Vernon (Ed.), *Speed of information-processing and intelligence* (pp. 239–270). Norwood, NJ: Ablex.
- KEATING, D. P., & MACLEAN, D. J. (1988). Reconstruction in cognitive development: A post-structuralist agenda. In P. B. Baltes, D. L. Featherman, & R. M. Lerner (Eds.), *Lifespan development and behavior* (Vol. 8, pp. 283–317). Hillsdale, NJ: Erlbaum.
- KUHN, D., AMSEL, E., & O’LOUGHLIN, M. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic Press.
- LEWIS, M. D. (1993). Emotion-cognition interactions in early infant development. *Cognition and Emotion*, 7, 145–170.
- LURIA, A. R. (1976). *Cognitive development: Its cultural and social foundations* (M. Lopez-Morrillas & L. Soltaroff, Trans.). Cambridge, MA: Harvard University Press.
- OATLEY, K., & JENKINS, J. M. (1992). Human emotions: Function and dysfunction. *Annual Review of Psychology*, 43, 55–85.
- ODEN, D., THOMPSON, R., & PREMACK, D. (1990). Infant chimpanzees spontaneously perceive both concrete and abstract same/different relations. *Child Development*, 61, 621–631.
- OLSON, D. R. (1977). From utterance to text: The bias of language in speech and writing. *Harvard Educational Review*, 47, 257–281.
- PIAGET, J. (1950). *The psychology of intelligence*. London: Routledge & Kegan Paul.

- PLOMIN, R., & THOMPSON, L. (1988). Life-span developmental behavior genetics. In P. B. Baltes, D. L. Featherman, & R. M. Lerner (Eds.), *Life-span development and behavior* (Vol. 8, pp. 1–31). Hillsdale, NJ: Erlbaum.
- SAVAGE-RUMBAUGH, S., McDONALD, K., SEVCIK, R., HOPKINS, W., & RUPERT, E. (1986). Spontaneous symbol acquisition and communicative use by pygmy chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: General*, 115, 211–235.
- SCRIBNER, S., & COLE, M. (1981). *The psychology of literacy*. Cambridge, MA: Harvard University Press.
- SHARP, D. W., COLE, M., & LAVE, J. (1979). Education and cognitive development: The evidence from experimental research. *Monographs of the Society for Research in Child Development*, 44 (Serial No. 178).
- STERNBERG, R. J. (1989). Domain-general versus domain-specificity: The life and impending death of a false dichotomy. *Merrill-Palmer Quarterly*, 35, 115–129.
- STERNBERG, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. New York: Cambridge University Press.
- STRINGER, C., & GAMBLE, C. (1993). *In search of the Neanderthals*. New York: Thames and Hudson.
- SUOMI, S. J. (1991). Early stress and adult emotional reactivity in rhesus monkeys. In *The childhood environment and adult disease* (Ciba Foundation Symposium 156). Chichester: Wiley.
- TERMAN, L. S. (1916). *The measurement of intelligence*. Boston: Houghton-Mifflin.
- TOMASELLO, M., KRUGER, A. C., & RATNER, H. H. (1993). Cultural learning. *Behavioral and Brain Sciences*, 16, 495–511.
- VIYGOTSKY, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- WERTSCH, J. (1985). *Vygotsky and the social formation of mind*. Cambridge, MA: Harvard University Press.

DANIEL P. KEATING

CORRELATION COEFFICIENT See STATISTICAL CONCEPTS.

CREATIVITY Of all the mental faculties, creativity is perhaps the one that in our culture most strongly attracts the lay imagination. Persons per-

ceived as creative—such as Picasso, Einstein, or Martha Graham—acquire almost superhuman reputations. As Robert Sternberg has shown, when people are asked to list characteristics associated with creativity, they tend to choose such traits as drive, perspicacity, decisionmaking skills, imagination, integrated intellect, and lack of conventionality (Sternberg, 1988, p. 128). Yet it would be difficult to get a broad consensus among social scientists as to the nature, or even the existence, of creativity as a qualitatively separate mental process. Some psychologists claim that so-called creative thinking is no more than very rapid rational thought that even computers can duplicate. Others believe that creativity is a social attribution that serves social purposes. Just as we need to believe that judges are fair, even when the evidence is to the contrary, because we need to believe in the rationality of our social environment, so we bring ourselves to believe that individuals who by luck were able to accomplish something out of the ordinary have a special trait—creativity—that makes them deserving of their success.

Despite the many ambiguities and difficulties involved in making sense of creativity, the topic is clearly of central importance to the understanding of what human beings are, and of what they can do. Creativity raises questions about the possibility of freedom in human action, about the dynamics of evolution, about the limitations of the IQ metric as a measure of intelligence. For these and many other reasons, the study of creativity is one of the most exciting fields in psychology.

From earliest historical times, people perceived a difference between what we now call “intelligence” and what we call “creativity.” The Greeks at the beginning of Western civilization attributed the ability to produce a new poem or song to the intervention of the Muse, a supernatural being whose inspiration was required to break out of the boundaries of normal thought. This connection between divine inspiration and the production of novelty is still present in the origins of the current term *creativity*, which in the late Renaissance began to be used as an analogy drawn between the divine creation of the natural world, and the artist’s ability to bring forth new shapes or sounds. For a long time, creativity was attributed only to artists, perhaps because until recently works of art were

the most remarkable examples of human artifice. Up through the Renaissance, artists had to have state-of-the-art knowledge of chemistry, geometry, and anatomy, as well as religion, philosophy, and literature. They produced the most impressive and sophisticated artifacts. This no longer being the case, in more recent times the popular attribution of creativity has tended to shift from the arts to the sciences, recognizing that the technological ingenuity of an Edison, or the theoretical scope of an Einstein's thought, is worthy of being designated creative.

THE PSYCHOLOGICAL STUDY OF CREATIVITY: A BRIEF HISTORY

Early Influences. Despite the intrinsic interest of the topic, creativity has not been a central concern in the history of psychology. Except for a few treatments of the subject by Sigmund Freud, who used a biographical reconstruction of Leonardo da Vinci's childhood to reflect on the sublimation involved in the curiosity and the relentless drive characteristic of creative individuals, and by Carl G. Jung, who was interested in the wellspring of creativity in the collective unconscious, the great majority of psychologists ignored the subject entirely. This neglect is understandable in light of the generally reductionistic tendency of the discipline during its first century of existence. The early psychophysical approach in Germany, the behaviorist hegemony in the United States, and the great commitment of effort to the measurement and study of intelligence narrowly defined all but excluded creativity from the conceptual vocabulary of the new science.

The origins of serious interest in the topic have been traced in part to the slow effects of the gestalt psychology of Max Wertheimer and Kurt Lewin, on the one hand, and the personality psychology of Gordon W. Allport and Henry A. Murray, on the other (MacKinnon, 1968, p. 435). The former demonstrated that complex interactions in perception and thought could be rigorously studied, and the latter demonstrated that individuals constructed their lives in terms of complex and sometimes original scripts. Some of the classic studies of this period include Ann Roe's investigation of creative physical scientists, which suggested that isolation from people in early childhood was a powerful motivator for their involvement with

abstract science (Roe, 1952), and her study of the personality of artists (Roe, 1946).

Definition of Creativity. By 1950, a general consensus had developed about how to define creativity, and the definition is still current today. Creativity is an attribute of ideas or products that (1) are original, or statistically infrequent, and therefore unpredictable, in a given culture; (2) are held to be valuable by the culture as a whole, or by a field of experts whose opinion is held to be legitimate by the culture; and (3) are carried on to a final, or at least to a useful, completion. It follows that creative individuals are those who come up with such ideas and products, the creative process is the one that results in them, and a creative environment is the one that fosters their production. Hence, creativity research has focused on the study of creative products, creative individuals, creative processes, and creative environments.

Guilford's Psychometric Approach. But sustained work on creativity, as indicated by citations on this topic in the *Psychological Abstracts*, did not begin until J. P. Guilford in 1950 entitled his presidential address to the American Psychological Association "Creativity." Guilford, who during World War II was charged by the U.S. Air Corps to develop tests that would select pilots who could deal with sudden emergencies, had become convinced by his findings that conventional intelligence, or *convergent thinking*, was a mental process relatively independent of originality, or *divergent thinking* (Guilford, 1956). A pilot who demonstrated superior intelligence in terms of abstract reasoning, memory, and ability to follow routine instructions might not be able to cope with new situations that required a reversal of normal procedures, or an entirely unprecedented response. Those who could do the latter were said to be divergent thinkers.

Under Guilford's influence, a geometric increase in the number of publications devoted to the study of creativity—or divergent thinking—took place. This spurt of interest was also helped by the national concern about falling behind the former Soviet Union in scientific achievements, precipitated by the first Soviet space probe launched on October 4, 1957, a fear that in turn was translated into financial support for research aimed at improving U.S. education, especially high-level scientific training, and thus included support of creativity research.

In the following decades the bulk of work in the area was aimed at developing reliable and valid creativity tests. This direction, modeled in large part on the previous history of IQ tests, investigated the psychometric characteristics of various measures that could be used to identify creativity in children, or predict creative behavior in adults. One of the main centers of research in this area has been the laboratory of E. P. Torrance (1963, 1987), who constructed a number of creativity tests for children, and has documented reasonably high correlations between early test scores and "creative" real-life performance decades later, such as starting a business, journal, or organization (Torrance, 1988).

The psychometric approach, however, has not been without its critics. For example, Howard Gruber, who has studied carefully Darwin's journals in an attempt to reconstruct the mental processes leading to the formulation of evolutionary theory (Gruber, 1981), has warned about the error of assuming that when children score high on a divergent-thinking test that takes a few minutes to complete, this has anything in common with the kinds of mental processes that over a lifetime lead to genuine creative contributions (Gruber, 1982, 1988).

An important landmark at midcentury was the work of Getzels and Jackson (1962), which demonstrated that IQ and creativity, which previously were often treated interchangeably, were really quite distinct abilities. Up to an IQ of about 120, creativity and IQ were, indeed, impossible to distinguish, but after that point higher IQ scores did not necessarily go with higher creativity test scores, or vice versa.

During this period a center for the study of creative behavior was also established at the State University of New York, Buffalo, where the *Journal of Creative Behavior* started publishing in 1967. One of the center's goals was to act as a bridge between academic research and practical application, especially in the world of business.

CURRENT DIRECTIONS IN THE STUDY OF CREATIVITY

Because of the complexity of the subject, it is impossible at this time to give a single coherent account of what is known about creativity, or even the main

direction of thought on the topic. Instead, we shall review some of the more notable current approaches, hoping in the process to illustrate the richness and variety of problems they present. Of course, placing a given research agenda in a single category also distorts the record to a certain extent, because some approaches could be listed under more than one heading. For instance, anyone who studies creative individuals will also be interested in the creative process, and perhaps in the effects of creative environments. Nevertheless, to facilitate the review, we shall discuss the main dimensions of creativity separately from each other.

Psychometric Approaches. The measurement of creative thinking through a variety of divergent thinking and other pencil-and-paper tests has continued to the present, and still has many strong advocates (Milgram, 1990; Hong & Milgram, 1991). In part this can be explained by the fact that testing has many immediate practical applications. Just as the popularity of the IQ test was due to its use in the selection of army recruits for World War I and later as a tool for stratifying students in the educational system, so interest in testing for creativity is fueled by special education policies that try to identify gifted and talented students for educational programs suited to their abilities. Yet while it is true that results obtained with tests purporting to measure creativity have reasonable reliability and seem to correlate with some sensible outcomes, it is fair to say that the testing approach has not advanced theory, and therefore has failed to capture the imagination of many scholars who try to understand the nature of the creative process.

Mathematical Simulations. Perhaps the newest direction in studying creativity is the one represented by mathematical modeling of novel thought processes (e.g., Findlay & Lumsden, 1988; Moneta, 1992), an approach that has been influenced by spreading activation theory (Anderson, 1983), and by sociobiological theory (Lumsden & Findlay, 1988). In Findlay and Lumsden's models, a discovery is defined as the formation of a new relationship between concepts in the subject's semantic network. An individual's creative potential is defined by an entropylike function that models the probability of occurrence of different mental outcomes in response to the same stimulus, thus reflecting Guilford's notions of mental

fluency and *flexibility*, the main components of divergent thinking. Mathematical models have some strengths and weaknesses opposite to those of the psychometric approaches to creativity. While they are elegant and wide-ranging, they are only heuristic devices that suggest hypotheses to be tested. Powerful as they may be, by themselves they do not add empirical substance to the understanding of creativity.

Computer Simulations. The computer simulation of creative problem-solving processes has been most vigorously pursued at Carnegie-Mellon University, in the laboratory of Herbert Simon. Here, the data involved in some important scientific breakthrough are fed into a computer, and then different heuristic programs, such as BACON, are used to achieve a solution. Thus, for instance, the program will reproduce Hans Krebs's steps in discovering how to synthesize urea *in vivo*, if the information available to Krebs is provided (Kulkarni & Simon, 1988); it will derive the basic classification of chemical substances from knowledge of their properties; or it will rediscover Kepler's Third Law in a few seconds if it has access to the relevant data on planetary distances and periods of revolution (Langley et al., 1987). While these feats of replicating creative discoveries are impressive, questions remain as to their relevance, given the fact that what differentiates creative solutions is precisely that they are *not* replications.

Secondary Analysis of Archival Data. An influential branch of creativity research involves the systematic collection and analysis of data about creative products and creative individuals in their historical contexts. For instance, Dean Simonton, the chief practitioner of this type of analysis, after collecting basic information on thousands of individuals from historical records in Europe, the United States, and China, has written extensively about age trends and productivity in the lives of creative individuals, about the social and political characteristics of particularly creative historical periods, and about shifts in the thematic content of creative works through time (Simonton, 1988, 1990).

Colin Martindale (1978, 1990) has focused particularly on the arts, and by analyzing samples from poems, paintings, and musical compositions representative of several successive centuries of Western history, has tried to describe the evolutionary dynamics

of creativity in the arts. These approaches, like the previous ones, are influenced by evolutionary theory, especially as construed in Donald Campbell's concepts of evolutionary epistemology and random variation (Campbell, 1974).

Creative Individuals. More than any other approach, the study of outstanding individuals—geniuses, prodigies, influential innovators—has been the staple of creativity research. Some of the themes have been extremely persistent: the Italian physician Cesare Lombroso discussed the relationship between genius and insanity as early as 1876 (Lombroso, 1910); the same question is still being debated (Andreasen, 1987; Ludwig, 1989; Prentky, 1980). Although certain psychopathologies, such as depression, substance abuse, and psychosis, seem to appear more frequently than expected among creative individuals, especially among those involved in the arts, the causal connections are by no means clear. It is very possible, for instance, that pathology plays no role whatsoever in the production of creativity, but is simply the result of the difficulty creative individuals encounter in having their ideas accepted.

The personality of creative individuals has also been extensively studied, by MacKinnon (1961, 1964), Barron (1972), and Guilford (1986), among others; in-depth studies of the lives and works of such persons have been conducted by Gruber (1981), Getzels and Csikszentmihalyi (1976), and Gardner (1988, 1993). Several common traits keep emerging from such studies. For instance, creative individuals are found to be very self-reliant, yet sensitive; they often display androgynous characteristics; they are often marginal to the culture in which they live; and they are often found to behave in a childlike manner.

However, other scholars have warned against taking the whole concept of "creative individuals" too seriously. It has been pointed out that it is often very difficult to attribute a creative idea or product to any one person; multiple discoveries are the rule rather than the exception, and the difference between a person who is rewarded with recognition and one who is ignored may not consist of anything more than chance (Ogburn, 1964; Merton, 1968). Attributionally oriented theorists would explain the existence of creative geniuses as due more to our need to worship greatness than to their inherent superiority (Brannigan, 1981).

The Creative Process. It is also difficult to separate myth from fact in dealing with the sequence of events preceding a discovery or creative achievement. The temptation has been strong—especially since the Romantic era—to embellish the narrative of the creative process with stories that perhaps ought to be true, but often are not. For instance, after Samuel Taylor Coleridge's poem *Kubla Khan* became very popular after its publication in 1816, he claimed that he had composed it in a flash of inspiration that came in a dream after he had drugged himself with opium. Recent research, however, has uncovered several early drafts of the same poem drafted in Coleridge's hand, suggesting that the opium story was a concession to the romantic expectations of his readers—and perhaps also to his own (Schneider, 1953).

Two major ways of looking at the creative process have been quite influential. According to the first of these approaches, psychoanalytically inspired investigators have tried to understand how the repressed contents of unfulfilled desires become transformed and expressed in works of art and science. Originally it was thought that creative ideation was under control of the unconscious, but gradually the concept of "regression at the service of the ego" has gained more support, and creativity is now seen to include an increasingly more reality-oriented direction (Kris, 1952; Kubie, 1958; Gedo, 1983).

Another recent approach involves applying insights from cognitive psychology to the understanding of the creative process. Examples include Sternberg's (1988) model integrating intelligence, intellectual style, and personality; Langley and Jones's (1988) computational model of scientific insight; and Getzels and Csikszentmihalyi's (1976) study of problem finding in artistic creativity. Other studies have focused on the effects of intrinsic motivation on creative production (Amabile, 1983, 1985) and the application of rational choice decision models borrowed from economic theory (Sternberg & Lubart, 1991; Rubenson & Runco, 1992).

Descriptions of the creative process usually differentiate three or four stages. The four stages made popular by the French mathematician Jacques-Salomon Hadamard (1954) begin with a period of *preparation*, during which the person becomes acquainted with the parameters of the problem and begins attempts at solution. This is followed by a period of *incubation*, lasting

anywhere from a few hours to several months, during which the problem is mulled over below the threshold of awareness. It is at this stage that psychoanalysts believe unconscious involvement to be at its highest, while cognitivists would describe this as a period when parallel rather than linear processing of mental associations takes place—which could well be the same phenomenon looked at from two different perspectives. Incubation is followed by a moment of *insight*, which provides a mental solution to the problem. But even the most compelling insight does not result in a creative contribution unless it is followed by a period of *evaluation* and *elaboration*, which is necessary to translate the personal vision into a product that could be shared and appreciated by others. While the stages of incubation and insight make the creative process seem to be very spontaneous and effortless, they are only part of the story. Preparation and elaboration are just as necessary, and it is because of them that it has been said that creativity is 90 percent perspiration and 10 percent inspiration.

One central question still unresolved is whether the creative process is the same in the sciences and the arts, or more generally, in the various domains of human activity (Feldman, 1980). While it seems that the very general four-stage model described above applies to all creative domains, it is also likely that each domain requires a somewhat different mix of, for instance, incubation versus elaboration, or preparation versus insight. Also one would expect that in artistic creativity the role of rational problem-solving steps would be less important than, say, in physics. Despite the obviousness of these issues, they have not been satisfactorily resolved.

Creative Environments. Historians (e.g., Toynbee, 1936) and anthropologists (e.g., Kroeber, 1944) have remarked on the fact that certain cultures at certain times produce an unusually high number of new ideas or artifacts. This implies that by modifying external conditions, the level of creativity in a given population might be increased. Among psychologists, Morris Stein (1974, 1991) has pioneered thinking about how institutional supports help or hinder creative processes. Teresa Amabile (1983, 1990) has focused on the social psychology of creativity, with special emphasis on educational and business settings. Recent writings have stressed the importance of look-

ing at the entire ecological context in which creativity takes place (Gruber, 1988; Harrington, 1990; John-Steiner, 1992).

Of the many social contexts that affect creativity, the one that has been studied most extensively is the family. It seems reasonable to expect that if any social institution has an effect on creativity, the early family environment must be it. Most of the research on this topic has been with gifted children, given the fact that creativity is so difficult to measure in early life (Albert & Runco, 1986; Bloom & Sosniak, 1981; Colangelo, 1988). The results are far from unambiguous. In the words of one researcher, "All that can be said about children destined for greatness is that they vary widely in their relationship with parents" (Tannenbaum, 1986, p. 46). It seems that very resilient children from dysfunctional families, or children from families that give extraordinary support and stimulation to their children, may both have a greater chance to make creative contributions than do children from average backgrounds (Albert, 1971; Simonton, 1987).

Part of the family's contribution to the development of creativity is simply the material support it may give or withhold. As Benjamin Bloom (1985) has shown in regard to the development of talented youth, small things, such as having books in the home, reading to children, chauffeuring them to and from places, and finding and paying for tutors, seem to be almost necessary to help youngsters to cultivate their skills to the point that they are ready to make a creative contribution. Yet here again, the evidence is far from clear. In their review of the childhood of hundreds of eminent—and often creative—persons, Goertzel and Goertzel (1962) have reported that geniuses have a way of disconfirming expectations and achieving great things despite very adverse early conditions.

OTHER APPROACHES

If one wishes to know all there is to know about creativity, it may not be wise to restrict oneself to what has been written under that rubric. The concepts of genius, talent, and giftedness all overlap in some ways with creativity. Innovation, entrepreneurship, and scientific discovery are some of the other names that scholars have given to the same phenomenon, depending on whether they were trained as sociologists,

economists, or historians of science (Wehner, et al., 1991). It is for this reason that the Hungarian psychologist Istvan Magyari-Beck (1990) has argued that any single existing discipline is too biased to accommodate the knowledge required to understand creativity, and that a new discipline, "creatology," is required fully to comprehend it. Whether such a discipline will materialize or not, it seems clear that a closer integration of the various approaches to the topic is called for.

In the meantime, however, creativity has been embraced by both business executives who hope to increase the innovativeness of their employees and the profitability of their companies and by various therapeutic movements that aim at liberating their clients from humdrum and uninteresting lives (*Utne Reader*, 1992). In some large corporations, such as Motorola, the emphasis on creativity training is a major concern. Many consultants and training programs specialize in enhancing creativity, and while there is little evidence that these interventions are actually beneficial, clients often report satisfaction with them.

THE SYSTEMS VIEW OF CREATIVITY

It seems to be the case that the phenomenon of creativity is best grasped not as something that applies to a person, a process, or a product, but as something that results from an *interaction* among these (and many other) factors. To understand creativity it may be necessary to view it as a property of a system made up of three interrelated components: a *person* who makes changes in the contents of a *domain* that are acceptable to a *field* (Csikszentmihalyi, 1988, 1990). A domain is any symbolic system, such as art, music, mathematics, medicine, basketball, a way of manufacturing widgets, or a way of programming computers. Domains form part of the broader culture. Fields are constituted by individuals who have the right to admit changes into the content of domains: They include art critics, collectors, patrons, and museum curators in the domain of art; conductors, critics, and music teachers in the domain of music; teachers, journal editors, and textbook writers in the domain of mathematics, and so on. Fields are part of the larger social system.

Such a model is required because it is impossible for a person to be creative except by operating within

a domain, and it is impossible to assess the creativity of a person's contribution without relying on the judgment of a field. For instance, the creative achievements of Renaissance Florence were not due to the fact that a great number of creative individuals were born in that town in the later part of the fourteenth century. Rather, they were due to the confluence of two other factors: The domain of art, and especially architecture, was suddenly enriched by the discoveries of previously buried and forgotten classical ruins in Rome and elsewhere; and a vigorous field emerged simultaneously, consisting of wealthy bankers and powerful nobles determined to transform Florence into the most beautiful city in Europe. It was this field that stimulated able young men, who otherwise would have gone into business or the professions, to try their hand at building cathedrals or painting frescoes. And these young men, inspired by the rediscovered rules of classical art and the interest of their contemporaries, were able to create works of a novelty and excellence that still astound us.

This system model works in a manner that is very similar to classical evolutionary models. Evolution happens when there is variation among individuals, when the most adaptive variations are selected by the environment, and when there is a way to preserve the best variations and to transmit them down to a new generation. In the systems model, the person provides the novel variation, the field does the selection, and the domain takes care of preserving and transmitting the selected variations. Each new generation encounters a domain enriched by the contributions of the previous generation, and tries to enrich it in its turn; thus the history of creativity consists of an ascending spiral in which individuals, fields, and domains interact and jointly produce new forms.

The systems model is different from earlier psychological approaches that recognized the importance of social and cultural contexts in helping and hindering creativity, but still assumed that creativity resided in the individual. The systems model claims that creativity is jointly constituted by these three separate agents. The individual contributes novel thoughts or products, but these cannot be called creative until they are added to the domain.

Looking at the issue this way brings a whole range of new questions into view. For instance, the impor-

tance of domains suggests the following questions: What forms of symbolic coding make it easier for individuals to innovate? How can domains be changed to make innovation more likely? Are domains that are clearly organized internally more or less likely to produce creativity? Are domains that are closer to the core values of the culture easier or more difficult to change?

Similarly, many questions concerning the field could help us understand better how creativity happens. What are the best incentives that the field can provide to stimulate individuals to innovate? Is it better to have a field that is very selective, or one that is quite forgiving? Is it good for the field to have exclusive power over the domain? What happens to creativity when the field becomes too dependent on political, or financial, interests?

Such questions, which in the past would have had only a marginal interest to students of creativity, are highlighted as being of central importance by the systems model. Of course, the process by which individuals grasp new ideas, and then bring them to fruition, will still remain a fundamental question. However, it may be necessary to realize that creativity is not something that happens inside a person's head, but something that develops in a complex interchange within a network of symbols and actors. It is by following this direction that the many seemingly unrelated aspects of this fascinating domain may be best integrated and comprehended.

(See also: GENIUS.)

BIBLIOGRAPHY

- ALBERT, R. S., & RUNCO, M. A. (1986). The achievement of eminence: a model based on a longitudinal study of exceptionally gifted boys and their families. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 332–360). Cambridge: Cambridge University Press.
- AMABILE, T. M. (1983). *The social psychology of creativity*. New York: Springer Verlag.
- AMABILE, T. M. (1985). Motivation and creativity: Effects of motivational orientation on creative writers. *Journal of Personality and Social Psychology*, 45, 357–377.

- AMABILE, T. M. (1990). Within you, without you: The social psychology of creativity, and beyond. In M. A. Runco & R. S. Albert (Eds.), *Theories of creativity* (pp. 61–91). Newbury Park, CA: Sage.
- ANDERSON, J. R. (1983). *The architecture of cognition*. Cambridge, MA: Harvard University Press.
- ANDREASEN, N. C. (1987). Creativity and mental illness: prevalence rates in writers and their first-degree relatives. *American Journal of Psychiatry*, *144*, 1288–1292.
- BARRON, F. (1972). *Artists in the making*. New York: Seminar Press.
- BLOOM, B. S. (1985). *Developing talent in young children*. New York: Ballantine.
- BLOOM, B. S., & SOSNIAK, L. A. (1981). Talent development vs. schooling. *Educational Leadership*, *39*, 86–94.
- BRANNIGAN, A. (1981). *The social basis of scientific discoveries*. Cambridge: Cambridge University Press.
- CAMPBELL, D. T. (1974). Evolutionary epistemology. In P. A. Schilpp (Ed.), *The philosophy of Karl Popper*. La Salle: Open Court.
- COLANGELO, N. (1988). Families of gifted children: The next ten years. *Roeper Review*, *11*, 16–18.
- CSIKSZENTMIHALYI, M. (1988). Society, culture, person: A systems view of creativity. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 325–339). Cambridge: Cambridge University Press.
- CSIKSZENTMIHALYI, M. (1990). The domain of creativity. In M. A. Runco & R. S. Albert (Eds.), *Theories of creativity* (pp. 190–214). Newbury Park, CA: Sage.
- FELDMAN, D. H. (1980). *Beyond universals in cognitive development*. Norwood, NJ: Ablex.
- FINDLAY, C. S., & LUMDSEN, C. J. (1988). The creative mind: Towards an evolutionary theory of discovery and innovation. *Journal of Social and Biological Structures*, *11*, 3–55.
- GARDNER, H. (1988). Creative lives and creative works: a synthetic scientific approach. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 298–324). Cambridge: Cambridge University Press.
- GARDNER, H. (1993). *The creators of the modern era*. New York: Basic Books.
- GEDO, J. (1983). *Portraits of the artist*. New York: Guilford Press.
- GETZELS, J. W., & CSIKSZENTMIHALYI, M. (1976). *The creative vision*. New York: Wiley.
- GETZELS, J. W., & JACKSON, P. (1962). *Creativity and intelligence*. New York: Wiley.
- GOERTZEL, T. G., & GOERTZEL, M. G. (1962). *Cradles of eminence*. Boston: Little, Brown.
- GRUBER, H. (1981). *Darwin on man: A psychological study of scientific creativity*. Chicago: University of Chicago Press.
- GRUBER, H. (1982). On the hypothesized relation between giftedness and creativity. In D. H. Feldman (Ed.), *Developmental approaches to giftedness and creativity* (pp. 7–30). San Francisco: Jossey-Bass.
- GRUBER, H. (1988). The evolving systems approach to creative work. *Creativity Research Journal*, *1*, 27–51.
- GUILFORD, J. P. (1956). Structure of intellect. *Psychological Bulletin*, *53*, 267–293.
- GUILFORD, J. P. (1959). *Personality*. New York: McGraw-Hill.
- HADAMARD, J.-S. (1954). *An essay on the psychology of invention in the mathematical field*. New York: Dover.
- HARRINGTON, D. M. (1990). The ecology of human creativity: a psychological perspective. In M. A. Runco & R. S. Albert (Eds.), *Theories of creativity* (pp. 143–169). Newbury Park, CA: Sage.
- HONG, E. S., & MILGRAM, R. A. (1991). Original thinking in preschool children: a validation of ideational fluency measures. *Creativity Research Journal*, *4*, 253–60.
- KROEBER, A. L. (1944). *Configurations of culture growth*. Berkeley: University of California Press.
- JOHN-STEINER, V. (1992). Creative lives, creative tensions. *Creativity Research Journal*, *5*, 99–108.
- KRIS, E. (1952). *Psychoanalytic explorations in art*. New York: International Universities Press.
- KUBIE, L. S. (1958). *Neurotic distortions in the creative process*. Lawrence: University of Kansas Press.
- KULKARNI, D., & SIMON, H. A. (1988). The process of scientific discovery: The strategy of experimentation. *Cognitive Science*, *12*, 139–75.
- LANGLEY, P., SIMON, H. A., BRADSHAW, G. L., & ZYTKOW, J. M. (1987). *Scientific discovery*. Cambridge, MA: MIT Press.
- LOMBROSO, C. (1910). *The man of genius*. London: Walter Scott.
- LUDWIG, A. (1989). Reflections on creativity and madness. *American Journal of Psychotherapy*, *43*, 4–14.
- LUMDSEN, C. J., & FINDLAY, C. S. (1988). Evolution of the creative mind. *Creativity Research Journal*, *1*, 75–91.
- MACKINNON, D. W. (ED.). (1961). *The creative person*. Berkeley: University of California, Institute for Personality Assessment Research.
- MACKINNON, D. W. (1964). The creativity of architects. In C. W. Taylor (Ed.), *Widening horizons in creativity*. New York: Wiley.

- MACKINNON, D. W. (1968). Creativity: Psychological aspects. In D. L. Sills (Ed.), *International Encyclopedia of the Social Sciences* (pp. 435–442). New York: Macmillan.
- MAGYARI-BECK, I. (1990). An introduction to the framework of creatology. *Journal of Creative Behavior*, 24, 151–160.
- MARTINDALE, C. (1978). The evolution of English poetry. *Poetics*, 7, 231–248.
- MARTINDALE, C. (1990). *The clockwork muse: The predictability of artistic change*. New York: Basic Books.
- MERTON, R. K. (1968). The Matthew effect in science. *Science*, 159, 56–63.
- MILGRAM, R. M. (1990). Creativity: an idea whose time has come and gone? In M. A. Runco & R. S. Albert (Eds.), *Theories of creativity* (pp. 215–233). Newbury Park, CA: Sage.
- MONETA, G. (1992). A model of scientists' creative potential: The matching of cognitive structure and domain structure. *Philosophical Psychology*, 5.
- OGBURN, W. F. (1926[1964]). The great man versus social forces. In O. D. Duncan (Ed.), *William F. Ogburn on culture and social change: Selected papers*. Chicago: University of Chicago Press, 44–61.
- PRENTKY, R. A. (1980). *Creativity and psychopathology: a neurocognitive perspective*. New York: Praeger.
- ROE, A. (1946). The personality of artists. *Educational and Personality Measurement*, 6, 401–408.
- ROE, A. (1952). *The making of a scientist*. New York: Dodd, Mead.
- RUBENSON, D. L., & RUNCO, M. A. (1992). The psychoeconomic approach to creativity. *New Ideas in Psychology*, 10, 131–147.
- SCHNEIDER, E. (1953). *Coleridge, opium, and Kubla Khan*. Chicago: University of Chicago Press.
- SIMONTON, D. K. (1987). Developmental antecedents of achieved eminence. *Annals of Child Development*, 5, 131–169.
- SIMONTON, D. K. (1988). *Scientific genius*. Cambridge: Cambridge University Press.
- SIMONTON, D. K. (1990). *Psychology, science, and history*. New Haven, CT: Yale University Press.
- STEIN, M. I. (1974). *Stimulating creativity*, Vol 1, *Individual procedures*. Orlando, FL: Academic.
- STEIN, M. I. (1991). On the sociohistorical context of creativity programs. *Creativity Research Journal*, 4, 294–300.
- STERNBERG, R. J. (1988). A three-facet model of creativity. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 125–147). Cambridge: Cambridge University Press.
- STERNBERG, R. J., & LUBART, T. I. (1991). An investment theory of creativity and its development. *Human Development*, 34, 1–31.
- TANNENBAUM, A. J. (1986). In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 21–52). Cambridge: Cambridge University Press.
- TORRANCE, E. P. (1963). *Education and the creative potential*. Minneapolis: University of Minnesota Press.
- TORRANCE, E. P. (1987). *The blazing drive: The creative potential*. Buffalo, NY: Bearly Ltd.
- TORRANCE, E. P. (1988). Creativity as manifest in testing. In R. J. Sternberg (Ed.), *The nature of creativity* (pp. 43–75). Cambridge: Cambridge University Press.
- TOYNBEE, A. J. (1936–1954). *A study of history* (10 vols.). Oxford: Oxford University Press.
- UTNE READER. (1992, March-April). Are you creative? *Utne Reader*, 52–69.
- WEHNER, L., CSIKSZENTMIHALYI, M., & MAGYARI-BECK, I. (1991). Current approaches used in studying creativity. *Creativity Research Journal*, 4, 261–271.

MIHALY CSIKSZENTMIHALYI

CRIME AND DELINQUENCY The etiology of delinquent and criminal behavior is an enigma that has been studied in depth. Delinquency refers to any behavior that is illegal for minors to engage in, including major infractions such as car theft and assault as well as more minor violations such as smoking or drinking under age. Criminality refers to behaviors that are illegal for adults to engage in. If we understood why individuals commit criminal acts, some repeatedly, then perhaps we would be able to design environmental interventions to reduce these behaviors in our society. In efforts to better understand the causes of delinquency and criminality, researchers have examined the environmental and genetic influences at work in molding these behaviors.

One of the frequently studied variables related to offense behaviors has been intelligence. Although an inverse relationship between intelligence and criminality often exists, some criminal acts require a fairly high level of intelligence for their execution (Wilson &

Herrnstein, 1985). As is usually the case in etiology, it is clear that an interaction of factors is responsible and that intelligence alone is not sufficient to answer the question of why some people become delinquent and/or criminal whereas others do not.

This entry addresses research on the relationship between intelligence and criminal behavior and investigates the roles of other variables, such as social class and school performance, that may affect that relationship. Finally, we synthesize research on the discrepancy between verbal versus performance IQ that has been associated with higher levels of delinquency. A number of studies have shown that delinquents tend to have higher performance IQ scores than verbal IQ scores. Although the magnitude of this difference may not be significant in all cases, the fact that performance scores are higher than verbal scores in all studies (Quay, 1987) is a pervasive finding that merits some attention.

IS THERE AN IQ–DELINQUENCY RELATIONSHIP?

A link between intelligence, as measured by IQ tests, and offense behavior has been well established in the research literature, with low IQ associated with delinquency and criminality. A large number of studies have shown that male delinquents tend to have lower IQ scores than nondelinquents and tend to have parents with lower IQ scores. One study (DiLalla, 1987) showed that both adult criminals and delinquents who continued to be criminal as adults had lower IQ scores than did delinquents who were not criminal as adults or adolescents who were neither delinquent nor criminal. The opposite effect of IQ also seems to hold, that of higher intelligence being a protective factor, Kandel and colleagues (1988) found that high-risk boys (boys with criminal fathers), who were not criminal themselves, had higher IQ scores than did high-risk boys who were criminal, or than low-risk boys, suggesting that high-risk boys with high IQs were able to avoid the pitfalls that led to criminality. It is unclear from this study, however, whether socioeconomic status (SES) or school performance played a role in later criminality. One study that included girls, by Ensminger, Kellam, and Rubin (1983), showed that low IQ

scores in girls were related to a decrease in delinquency.

The relationship between IQ and delinquency holds whether police records or self-report questionnaires are used to assess delinquency. This fact lends credence to the idea that the IQ–delinquency relationship is not simply a function of less intelligent criminals being caught. Moffitt and Silva (1988), using police records as well as self-reports and parent and teacher reports as indicators of problem behaviors, found that delinquents still had lower IQs than adolescents without these indicators of behavior problems.

The relationships between intelligence and violence and between intelligence and recidivism are more questionable. There have been several studies showing negative relationships and others showing no relationship. A possible confound in these studies is the fact that perpetrators of violent crimes are more likely to be arrested than perpetrators of nonviolent crimes. This may also hold for recidivists, who, having been arrested once, are more suspect when another crime is committed. The definition of recidivism is also problematic, with some studies using more than one crime as their cut-off and others using more than two offenses.

Even though delinquents tend to have lower IQ scores than do nondelinquents, they nonetheless do not have IQ scores that are significantly below the norm. Therefore, a huge deficiency in mental capacity is not the cause of their delinquent behaviors. It is clear, however, that there is some relationship between delinquency and intelligence, the intricacies of which have yet to be clarified. A number of studies have shown that IQ is a more important predictor of delinquency than is SES (for a review, see Hirschi & Hindelang, 1977; Wilson & Herrnstein, 1985). Even when controlling for SES, IQ remains significantly related to delinquency. IQ and SES both appear to be related to delinquency and criminality, which is not surprising because they are themselves so strongly related.

The next section explores other possible causes of delinquency and criminality that are related to intellectual characteristics and may augment the effects of decreased IQ or may be byproducts of lower IQ. At issue is whether IQ and other characteristics or vari-

ables, such as school performance or teacher attitudes, independently affect delinquency and criminality, or whether IQ directly affects these other characteristics, which in turn are causally related to delinquency and criminality.

INDIVIDUAL CHARACTERISTICS VERSUS SOCIETAL INSTITUTIONS

Two different viewpoints have been raised about the interactive effects of IQ and school performance on ensuing delinquency and criminality. One position, put forth by Hirschi and Hindelang (1977), suggests that school performance and attitudes toward school are responsible for causing later delinquency. They believe that children with low IQs tend to perform poorly in school, which in turn leads to negative attitudes toward school. This negative attitude prohibits proper identification with school as a societal mentor, and therefore these low-IQ youths do not develop proper societal mores and they turn instead to delinquency.

An alternative viewpoint is one suggested by Menard and Morse (1984). These authors point out that schools may react differently to children of different IQ levels. Hence, it is not that low-IQ children reject school because they perform badly and therefore reject societal values and become delinquent. Rather, Menard and Morse suggest that school personnel may react negatively to children with low IQs. This negative reaction leads to personnel not providing an equal environment for these children, possibly spending less time with them than with other students, giving them fewer rewards and privileges, and providing them with less access to stimulating materials. This in turn leads to increased negativity by the children and eventually to delinquency.

Menard and Morse (1984) tested both models with a sample of 257 ninth graders. They found that including IQ in their model did not improve the prediction of delinquency. The variables that were most predictive of nonserious self-reported delinquency were social labeling as a problem child by teachers, friends, and parents, and delinquent peer group association. The most predictive variables of serious delinquency were, primarily, delinquent peer group association, as well as low academic aptitude, gender (being male),

and social labeling. Strangely, Menard and Morse included an enormous number of variables, all of which should have been sufficient to explain the causes of delinquency quite well. However, they were able to explain only a small portion of what causes delinquency. This means that either there was a huge measurement error or they were missing the key ingredients for predicting delinquency. One of the variables not included in their study was a family background of criminality, which has been shown to be an important variable for explaining adult criminality and has been suggested to be important for delinquency as well (DiLalla & Gottesman, 1989). Another important variable not included was child personality (see Wilson & Herrnstein, 1985).

PERFORMANCE VERSUS VERBAL INTELLIGENCE

In 1958 Wechsler reported that the most salient characteristic of the intelligence test performance of delinquents was a pattern in which performance scores were significantly higher than verbal scores on the Wechsler-Bellevue intelligence test (a phenomenon notated as $P > V$). This statement served as the impetus in generating a large body of research that further explored the relationship between performance and verbal scores in adolescent and adult offenders. Although a few studies have failed to confirm the finding that delinquents score significantly higher on the performance scale than on the verbal scale, countless studies have found evidence to support its existence.

Although the $P > V$ relationship has been well documented in the literature, psychologists argue as to the utility of this knowledge. Much of this debate stems from the magnitude of discrepancy found in the literature. In a discussion of the $P > V$ literature, Culbertson, Feral, and Gabby (1989) stated that the performance minus verbal difference found in studies ranges from 5.6 to 15 points. Given that the standard deviation of the WAIS and WISC-R intelligence tests is fifteen points, a difference of eight to ten points, which is commonly found, is of marginal practical significance. The degree of performance versus verbal discrepancy, however, is useful in predicting the degree of violence an offender exhibits (Walsh, Petee, & Beyer, 1987). In a more fine-grained analysis, Quay

(1987) analyzed the verbal scores from studies that explored P versus V . He reported that most studies found verbal IQ scores of delinquents to be about ten to twelve points below the expectation for the general population, a discrepancy large enough to indicate a possible verbal disadvantage.

In general, recidivists have been found to score lower on intelligence tests than one-time offenders. Haynes and Bensch (1981, 1983) found that both male and female recidivists were more likely to exhibit $P > V$ than were nonrecidivists. However, within the group of recidivists, the data did not indicate that the $P > V$ discrepancy is increasingly likely with additional adjudications. Haynes and Bensch therefore concluded that although differential rates of $P > V$ can distinguish groups of one-time offenders from recidivists, recidivists with few offenses cannot be distinguished from those with many offenses.

Low intelligence also has been linked to violent crime. Delinquents who score in the dull-normal range on verbal intelligence have been shown to have the highest scores on violent crime ratings. In contrast, bright-normal delinquents seem to have the lowest violent-crime ratings. When comparing performance and verbal scale intelligence, another pattern emerges. Walsh, Petee, and Beyer (1987) divided male delinquents into three groups: $P > V$ (imbalanced); $P < V$ (imbalanced); and $P = V$. They found that delinquents who were $V > P$ or $P > V$ were significantly more violent than offenders who were $P = V$. It would be interesting to know whether either of the imbalanced groups was more violent than the other, but this information was not presented. The $P = V$ subjects comprised 55.6 percent of the delinquent population, whereas $V > P$ and $P > V$ represented 11.3 and 33.8 percent of the population, respectively. Therefore, although scale-imbalanced delinquents ($P > V$ or $V > P$) are less common than $V = P$ offenders, they are more often involved in violent crime.

CONCLUSIONS

Given the large number of studies supporting a relationship between intelligence and criminal behavior, it is clear that somehow these two factors are indeed related. What remains unclear is exactly how they affect each other. It is most likely that low intelligence

places children and adults at risk, making them vulnerable to environmental insults. Poor parenting, for instance, may have a greater influence on a child with low intelligence, who has few personal resources to make up for the lack of parental guidance.

It seems clear also that school performance figures importantly in the equation. The direction of effects is not easily deciphered, however. Poor personal attitudes toward school may lead to poor school performance as well as poor performance on IQ tests, and these attitudes may also lead to delinquency (Quay, 1987; Rutter & Giller, 1984). Thus, attitude may be a third variable linking intelligence, school performance, and delinquency. In a related vein, children with high verbal abilities are likely to perform well in school, which may increase their likelihood of enjoying school and bonding with school-enforced values, and therefore may lead to greater law-abiding behavior.

In sum, this research has important implications for policies aimed at reducing delinquency and criminality. The fact that intelligence is related to offense behavior does not mean that nothing can be done about such behavior. One potentially useful framework for intervening is within the schools. Sessions designed to train teachers not to ignore, demean, or negatively label low-IQ students should ultimately improve the student-teacher relationship. In-class and extracurricular school activities designed to involve low-IQ students would help make school a more positive experience for these students. Such interventions need to be tried in an effort to help reduce or eliminate the IQ-delinquency-criminality relationship.

(See also: CRIMINALITY; ENVIRONMENT.)

BIBLIOGRAPHY

- CLONINGER, C. R., & GOTTESMAN, I. I. (1987). Genetic and environmental factors in antisocial behavior disorders. In S. Mednick & T. Moffitt (Eds.), *Biosocial bases of antisocial behavior*. Cambridge: Cambridge University Press.
- CULBERTON, F. M., FERAL, C. H., & GABBY, S. (1989). Pattern analysis of Wechsler Intelligence Scale for Children-Revised profiles of delinquent boys. *Journal of Clinical Psychology*, 45, 651-660.

- DiLALLA, L. F. (1987). Forecasting transitions from delinquency to criminality. Unpublished doctoral dissertation, University of Virginia, Charlottesville, VA.
- DiLALLA, L. F., & GOTTESMAN, I. I. (1989). Heterogeneity of causes for delinquency and criminality: Lifespan perspectives. *Development and Psychopathology, 1*, 339–349.
- DUNHAM, R. E. (1954). Factors related to recidivism in adults. *Journal of Social Psychology, 39*, 77–91.
- ENSMINGER, M. E., KELLAM, S. G., & RUBIN, B. R. (1983). School and family origins of delinquency: Comparisons by sex. In K. T. Van Dusen & S. A. Mednick (Eds.), *Prospective studies of crime and delinquency*. Boston: Kluwer-Nijhoff Publishing Co.
- FARRINGTON, D. P. (1975). The family backgrounds of aggressive youths. In L. Hersov et al. (Eds.), *Aggressive and antisocial behavior in childhood and adolescence*. Oxford: Pergamon.
- FARRINGTON, D. P., & WEST, D. J. (1971). A comparison between early delinquents and young aggressives. *British Journal of Criminology, 11*, 341–358.
- GANZER, V. J., & SARASON, I. G. (1973). Variables associated with recidivism among juvenile delinquents. *Journal of Consulting and Clinical Psychology, 40*, 1–5.
- GORDON, R. A. (1975). Crime and cognition: An evolutionary perspective. *Proceedings of the II International Symposium on Criminology*. São Paulo, Brazil: International Center for Biological and Medico-Forensic Criminology.
- GORDON, R. A. (1987). SES versus IQ in the race-IQ-delinquency model. *International Journal of Sociology and Social Policy, 7*, 30–96.
- GRACE, W. C., & SWEENEY, M. E. (1986). Comparisons of the $P > V$ sign on the WISC-R and WAIS-R in delinquent males. *Journal of Clinical Psychology, 42*, 173–176.
- GUERTIN, W. H., LADD, C. E., FRANK, G. H., RABIN, A. E., & HIESTER, D. S. (1966). Research with the Wechsler Intelligence Scale for Adults, 1960–1965. *Psychology Reports, 65*, 385–409.
- HAYNES, J. P., & BENSCH, M. (1981). The $P > V$ sign on the WISC-R and recidivism in delinquents. *Journal of Consulting and Clinical Psychology, 49*, 480–481.
- HAYNES, J. P., & BENSCH, M. (1983). Female delinquent recidivism and the $P > V$ sign on the WISC-R. *Journal of Clinical Psychology, 39*, 141–142.
- HIRSCHI, T., & HINDELANG, M. J. (1977). Intelligence and delinquency: A revisionist review. *American Sociological Review, 42*, 571–587.
- HUTCHINGS, B., & MEDNICK, S. A. (1975). Registered criminality in the adoptive and biological parents of registered male criminal adoptees. In R. R. Fieve, D. Rosenthal, & H. Brill (Eds.), *Genetic research in psychiatry*. Baltimore: Johns Hopkins University Press.
- KANDEL, E., MEDNICK, S., KIRKEGAARD-SORENSEN, L., HUTCHINGS, B., KNOP, J., ROSENBERG, R., & SCHULSINGER, F. (1988). IQ as a protective factor for subjects at high risk for antisocial behavior. *Journal of Consulting and Clinical Psychology, 56*, 224–226.
- KOLLER, K. M., & GOSDEN, S. D. (1980). Recidivists: Their past and families compared with first time only offenders. *Australian and New Zealand Journal of Criminology, 13*, 117–123.
- LEWIS, D. O., SHELLEY, S. S., PINCUS, J. H., & GLASER, G. H. (1979). Violent juvenile delinquents: Psychiatric, neurological, psychological, and abuse factors. *Journal of American Academy of Child Psychiatry, 18*, 307–319.
- LIPSITT, P. D., BULKA, S. L., & LIPSITT, L. P. (1990). Early intelligence scores and subsequent delinquency: A prospective study. *The American Journal of Family Therapy, 18*, 197–208.
- MCGUFFIN, P., & GOTTESMAN, I. I. (1985). Genetic influences on normal and abnormal development. In M. Rutter & L. Hersov (Eds.), *Child and adolescent psychiatry: Modern approaches*, 2nd ed.—Oxford: Blackwell.
- MENARD, S., & MORSE, B. J. (1984). A structuralist critique of the IQ–delinquency hypothesis: Theory and evidence. *American Journal of Sociology, 89*, 1347–1378.
- MOFFITT, T. E., GABRIELLI, W., & MEDNICK, S. A. (1981). Socioeconomic status, IQ, and delinquency. *Journal of Abnormal Psychology, 90*, 152–156.
- MOFFITT, T. E., & SILVA, P. A. (1988). IQ and delinquency: A direct test of the differential detection hypothesis. *Journal of Abnormal Psychology, 97*, 330–333.
- OSBORNE, S. G., & WEST, D. J. (1979). Conviction records of fathers and sons compared. *British Journal of Criminology, 19*, 120–133.
- PFEFFER, C. R., PLUTCHIK, R., & MITZRUCHI, M. S. (1983). Predictors of assaultiveness in latency age children. *American Journal of Pediatrics, 140*, 31–35.
- QUAY, H. C. (1987). Intelligence. In H. C. Quay (Ed.), *Handbook of juvenile delinquency*. NY: Wiley.
- REICHEL, H. (1987). The intelligence–criminality relationship: A critical review. *Reports from the Department of Psychology, The University of Stockholm*.
- RUTTER, M., & GILLER, H. (1984). *Juvenile delinquency: Trends and perspectives*. NY: Guilford Press.

- WALSH, A. (1987). Cognitive functioning and delinquency: Property versus violent offenses. *International Journal of Offender Therapy and Comparative Criminology*, 31, 285–289.
- WALSH, A., PETEE, T. A., & BEYER, J. A. (1987). Intellectual imbalance and delinquency: Comparing high verbal and high performance IQ delinquents. *Criminal Justice and Behavior*, 14, 370–379.
- WECHSLER, D. (1958). *The measurement and appraisal of adult intelligence*, 4th ed.—Baltimore: Williams & Wilkins.
- WEST, D. J., & FARRINGTON, D. P. (1973). *Who becomes delinquent?* London: Heinemann Educational Books.
- WILSON, J. Q., & HERRNSTEIN, R. J. (1985). *Crime and human nature*. NY: Simon & Schuster.

LISABETH F. DiLALLA
LORRIE A. NIELSON

CRIMINALITY Historically, little differentiation was made among criminals, mentally ill people, and those of subnormal intelligence (Santamour & West, 1982). All were seen as abnormal and all were segregated from the rest of society. Around the end of the nineteenth century some commentators began to believe that mental retardation predisposed people to engage in criminal behavior (Brown & Courtless, 1971; Fernald, 1909; Guidry et al., 1978). In fact, Fernald believed that intelligence tests should be used to identify intellectually impaired individuals, whom he considered potential criminals, for lifelong segregation. Scherenberger (1982) noted that during that same time intellectually impaired women were particularly likely to be institutionalized because they were considered more likely to become prostitutes than women with high intelligence. Rooted in such history, a number of questions persist about the relationship between intelligence and criminality, and the way in which intellectually impaired people are treated in the criminal justice system.

INTELLECTUAL IMPAIRMENT AMONG OFFENDERS

Research indicates that approximately 10 percent of offenders are intellectually impaired or have levels of intelligence that are far below normal, compared with 3 percent in the general population of the United States. Brown and Courtless (1971) found considerable

variation in percentages of intellectually impaired offenders when different geographical regions of the United States were considered separately (i.e., East South Central (24.3%), West South Central (20.6%), Pacific (5.4%), Mountain (2.6%).

Brown and Courtless (1971) found that 6 percent of incarcerated women were mentally retarded. Santamour and West (1982), however, found a range of from 2.5 percent to 6.1 percent, depending on where the sampling was done.

A great deal of attention has been paid to whether people with low levels of intelligence are more prone to commit criminal acts than those with higher levels of intelligence. Wilson and Herrnstein (1985) concluded that criminals have an average IQ of 92 and they believe that this low IQ indicates that criminals cannot think past short-term horizons or may be unable to understand society's rules or the consequences of their actions. However, the fact that people in prison have lower levels of intelligence, in general, than people in the general population does not prove that having a low level of intelligence causes one to be a criminal. Rather, it could be that the criminals who are more likely to be caught are not as intelligent as those criminals who are not caught. Further, prison inmates may not be particularly motivated to take IQ tests; therefore, the validity of the tests for inmates must be questioned.

Hirschi and Hindelang (1977) note that delinquents score consistently lower on standard intelligence tests than do nondelinquents. They report that this relationship exists independent of socioeconomic class; in other words, high-socioeconomic-class delinquents have been found to have lower IQs than high-socioeconomic-class nondelinquents and low-socioeconomic-class delinquents have been found to have lower IQs than low-socioeconomic-class nondelinquents. These authors conclude that IQ is a better predictor of delinquency than socioeconomic class or race.

RELATIONSHIP BETWEEN TYPE OF CRIMES COMMITTED AND INTELLIGENCE

Although level of education is not a direct indication of intelligence, some researchers have investigated the relationship between education and intelligence.

Roundtree and Faily (1980) found that incarcerated women with ten to twelve years of education committed more acts of aggression than those with thirteen to sixteen years of education, seven to nine years of education, and zero to six years of education, respectively. Likewise, women in their sample with an IQ between 39–75 committed more rule violations than did women with an IQ of 110–123, 76–100, and 101–108, respectively.

Some studies have paid specific attention to the types of crimes for which intellectually impaired offenders have been found guilty. Shapiro (1986) reports that 33 percent of intellectually impaired offenders were in prison for sex offenses. Of those studied by Shapiro, 83 percent had been convicted of stealing at some time in their lives, and very few had convictions for vandalism. There was also a relatively high incidence of violence, often against the mother, in the home for this sample. Denkowski and Denkowski (1984) state that intellectually impaired adolescent offenders “recurrently engage in such acts as physical and verbal aggression, property destruction, noncompliance, lying, stealing and running away” (p. 13). Steiner (1984) found that 63 percent of the intellectually impaired offenders in his sample had been imprisoned for property crimes, 23 percent for crimes against the person, and only 15 percent for sexual offenses.

Although much of the research investigating the relationship between intelligence and creativity shows correlations and lacks any explanatory theories, Lefkowitz et al. (1977) suggest that the frustrations of poor academic achievement and limited intellectual capacity may instigate violent antisocial behavior. Kandel and Mednick (1988) investigated the possible role that intelligence might play in influencing whether children from various backgrounds pursue a life of crime. Children who were at high risk for leading a life of crime (as indicated by their background) but who resisted criminal involvement had higher intelligence scores than those children who did end up committing crimes.

PATTERNS OF CRIMINALITY ASSOCIATED WITH INTELLECTUAL DEFICIT OR EXCESS

Some clinicians believe that people who have been diagnosed as psychopaths (antisocial personality dis-

order) may have higher than average levels of intelligence. As the following information shows, the research results have been mixed. Some researchers found that levels of intelligence did not differ significantly between samples of psychopaths and non-psychopaths (e.g., Sutker & Allain, 1987). Other researchers found that intelligence was an important factor in differentiating people who commit violent offenses from those whose offenses are nonviolent (e.g., Holland, Beckett, & Levi, 1981).

Heilbrun (1979) proposed that a combination of high psychopathy and low intelligence might represent an effective basis for predicting dangerous behavior. This conclusion has been supported by Heilbrun and other researchers (Wilson & Herrnstein, 1985). In 1982, Heilbrun tried to determine whether cognitive functioning interacts with psychopathy and intelligence to predict criminal behavior. His findings revealed that more intelligent criminals displayed superior cognitive control independent of their history of violent or nonviolent crime. Likewise, low-IQ psychopaths demonstrated the poorest impulse control, and their crimes were also more likely to include murder and rape.

Heilbrun (1990) attempted to discriminate offenders within a violent criminal sample on the basis of IQ and antisociality, which he combined to produce a “dangerousness index.” Results indicated that higher scores on the dangerousness index (i.e., low levels of intelligence and high levels of antisociality) correspond to more violent criminals. Heilbrun concluded that high antisociality and low IQ may lead to serious violence because situations become complicated and frustrating for the cognitively limited, antisocial man.

Aside from studies about the relationship between intellectual functioning and psychopathy or the commission of violent offenses, researchers have also been interested in investigating the relationship between intelligence and other types of criminality. Tammany, Evans, and Barrett (1990) studied the performance of felony offenders on intellectual-ability measures. Drug offenders scored highest on the intelligence measure. Property offenders had relatively low scores on intelligence suggesting, according to the authors, a simplistic, concrete type of thinking. The level of intelligence of people in the most violent offender group did not differ in any way from other groups.

Some attention has been paid to levels of intelligence of sex offenders. Quinsey, Arnold, and Pruesse (1980) found a higher verbal IQ for murderers than for sex offenders or arsonists. Other studies indicate that sex offenders who had committed crimes against children (pedophiles) are usually similar to the general population, with only a slight skew toward the lower end of the intelligence scale. In another study, Hucker et al. (1986) found that pedophiles collectively scored significantly lower on general intelligence and performance intelligence, but not on verbal intelligence measures.

As the above information shows, there are relatively high percentages of low-intelligence offenders. Therefore, the next question is whether such offenders are treated differently from other offenders in the criminal justice system.

CRIMINAL JUSTICE TREATMENT OF INDIVIDUALS WITH LOW INTELLIGENCE

Garcia and Steele (1988) note that intellectually impaired offenders face special problems in the criminal justice system. Few programs help prevent these offenders from committing other offenses, and they seem to bounce back and forth between the criminal justice system and the mental health system (French, 1983).

In a survey, Schilt (1979) found that although judges, lawyers, and police officers in New York may have some understanding of intellectually impaired people, they generally do not know how to deal with these people in a professional manner.

Intellectually impaired people may not be capable of appreciating the rights that are available to people in the criminal justice system. For example, they may not understand their Miranda warnings (see Garcia & Steele, 1988) and they often waive their rights upon arrest and sign a confession without the advice of a lawyer (French, 1983). Hall, LaFave, and Israel (1969) found that 95 percent of the intellectually impaired offenders they studied either confessed to the crime they were accused of or pleaded guilty. Similarly, the issue of competency to stand trial can also pose special problems for the intellectually impaired offender. People who are found incompetent to stand trial are often

committed to an institution until such time as they become competent. This may never happen with the intellectually impaired offender, since mental retardation may be especially unresponsive to "treatment," even if it is available.

The pretrial phase also offers some unique problems for the intellectually impaired offender. The decision about whether an individual should be granted bail or bond is often based on factors such as stability, employment, community ties, family strength, prior record, and permanent residence (Morrow, 1976). It is unlikely that these qualities are going to be found in many intellectually impaired offenders, especially those who have been in state facilities and/or those who have been deinstitutionalized.

Brown and Courtless (1971) made the following interesting observations regarding intellectually impaired offenders with an IQ under 55: 7.7 percent were not represented by an attorney; 59 percent pleaded guilty; 40 percent of those pleading guilty waived their rights to a jury trial; for 80 percent, the convicting charge was the same as the original charge (evidencing no plea bargaining); 66 percent made incriminating statements or confessions; 78 percent were not given a pretrial psychological evaluation; the issue of competency to stand trial was not raised in 92 percent of the cases; and, in 88 percent of the cases no appeals were made. The authors concluded that intellectually impaired offenders are not being treated equitably by the criminal justice system. This may also help to explain why the intellectually impaired are overrepresented in the criminal justice system.

When the intellectually impaired offender is placed in a facility with more sophisticated or hardened criminals, victimization and/or exploitation can occur (Garcia & Steele, 1988). Intellectually impaired offenders tend to be incarcerated at a younger age than other offenders (Manne & Rosenthal, 1971). Once incarcerated, they tend to remain incarcerated longer than other offenders (Santamour & West, 1982). Garcia and Steele (1988) note that a partial explanation of this latter tendency may be the inability of intellectually impaired offenders to successfully complete and/or participate in programs necessary for parole, their crimes may be more serious in nature, and frequent rule infractions during incarceration may result in their losing "good time." Steiner (1984) also notes that

intellectually impaired offenders may not be aware of the programs available to them in prison.

In 1989 the United States Supreme Court failed to overturn the death sentence of an intellectually impaired man convicted of murder (*Penry v. Lynaugh*). Johnny Paul Penry had suffered brain damage during birth and, as an adult, had an IQ of between 50 and 63. In 1980 Penry was sentenced to death after being convicted of raping and stabbing a woman to death. The sentence was eventually appealed to the Supreme Court, where the majority held that it is not necessarily unconstitutional (i.e., cruel and unusual) to execute mentally retarded offenders unless they were so "profoundly or severely retarded and wholly lacking the capacity to appreciate the wrongfulness of their actions" (p. 2954).

CONCLUSION

Overall, there does seem to be a mild relationship between intelligence and criminality. On average, criminals have lower IQ scores than noncriminals. Whether it is only the less intelligent criminals that get caught is an issue that remains undetermined. Criminal behavior self-reports suggest that this is not the case, but the evidence is far from conclusive. White-collar crime, for instance, is an area that has received little attention in the literature on crime and intelligence. There also appear to be some tentative relationships between the type of crime committed and the degree or type of intellectual deficit displayed by a given criminal. These findings, however, are, on the whole, somewhat equivocal. Abundant evidence exists to show that intellectually impaired offenders are not treated fairly in the criminal justice system.

There are a number of cautions regarding the interpretation of the relationship between crime and intelligence that should be highlighted:

1. There are numerous well-known pitfalls inherent in intelligence tests (e.g., cultural bias) that may put the research results in question.
2. There are numerous ways of defining clinical behaviors such as psychopathy, mental retardation, dangerous offenders, et cetera. Lack of uniformity on the matter makes general statements regarding the nature of crime and intelligence very difficult.

3. Other variables may exist that could explain the lower level of intellectual functioning among offenders. For instance, it could be the case that the prison experience dulls or, at the very least, does not facilitate intellectual development. Similarly, given the high incidence of alcohol and drug-related offenses, it could be that substance abuse by offenders leads to lower intelligence test performance.

(See also: CRIME AND DELINQUENCY; ENVIRONMENT.)

BIBLIOGRAPHY

- BROWN, B., & COURTLESS, T. (1971). *The mentally retarded offender* (DHEW Publication No. HSM 72-90-39). Washington, DC: U.S. Government Printing Office.
- DENKOWSKI, G. C., & DENKOWSKI, K. M. (1984). A residential treatment model for MR adolescent offenders. *Hospital and Community Psychiatry, 35*, 279-281.
- EBERLY, D. A., & KUTASH, K. (1984). *Florida adolescent and child treatment survey: Second interim report on adolescent and child clients at state training schools*. Tallahassee: State of Florida.
- FERNALD, W. E. (1909). The imbecile with criminal instincts. *Journal of Psycho-Asthenics, 14*, 16-23.
- FRENCH, L. A. (1983). The mentally retarded and pseudo-retarded offender: A clinical/legal dilemma. *Federal Probation, 43*, 55-61.
- GARCIA, S. A., & STEELE, H. V. (1988). Mentally retarded offenders in the criminal justice and mental retardation service systems in Florida: Philosophical, placement, and treatment issues. *Arkansas Law Review, 41*, 809-891.
- GUIDRY, L. L., MCBRIDE, W. A., & WATERS, T. J. (1978). WAIS performance of forensic versus civilly committed mental retardates. *Psychological Reports, 43*, 723-726.
- HALL, K., LAFAYE, W., & ISRAEL, J. (1969). *Modern criminal procedure* (3rd ed.). St. Paul, MN: West Publishing Co.
- HASKINS, J., & FRIEL, C. (1982). *Project CAMIO* (Vols. 1-8). Huntsville, TX: Sam Houston State University.
- HEILBRUN, A. B., JR. (1979). Psychopathy and violent crime. *Journal of Consulting and Clinical Psychology, 47*, 509-516.
- HEILBRUN, A. B., JR. (1982). Cognitive models of criminal violence based upon intelligence and psychopathy levels. *Journal of Consulting and Clinical Psychology, 50*, 546-557.
- HEILBRUN, A. B., JR. (1990). Differentiation of death-row

- murderers and life-sentence murderers by antisociality and intelligence measures. *Journal of Personality Assessment*, 54, 617–627.
- HIRSCHI, T., & HINDELANG, M. J. (1977). Intelligence and delinquency. *American Sociological Review*, 42, 571–587.
- HOLLAND, T. R., BECKETT, G. E., & LEVI, M. (1981). Intelligence, personality, and criminal violence: A multivariate analysis. *Journal of Consulting and Clinical Psychology*, 49, 106–111.
- HUCKER, R., LANGEVIN, G., WORTZMAN, J., BAIN, L., HANDY, J., CHAMBERS, J., & WRIGHT, S. (1986). Neuropsychological impairment in pedophiles. *Canadian Journal of Behavioural Science*, 18, 440–448.
- KANDEL, E., & MEDNICK, S. (1988). IQ as a protective factor for subjects at high risk for antisocial behavior. *Journal of Consulting and Clinical Psychology*, 56, 224–226.
- LANGEVIN, R., PAITICH, D., FREEMAN, R., MANN, K., & HANDY, L. (1978). Personality characteristics and sexual anomalies in males. *Canadian Journal of Behavioural Science*, 10, 222–238.
- LEFKOWITZ, M., ERON, L., WALDER, L., & HUESMANN, L. (1977). *Growing up to be violent: A longitudinal study at the development of aggression*. New York: Pergamon Press.
- MANNE, S. H., & ROSENTHAL, D. (1971). IQ and age of first commitment of dangerous offenders. *Correctional Psychologist*, 4, 220–229.
- MORROW, C. (1976). A legal framework: An insider's perspective. In P. Browning (Ed.), *Rehabilitation and the retarded offender*. Springfield, IL: Charles C Thomas.
- PENRY v. LYNNAUGH, 109 S.Ct. 2934 (1989).
- QUINSEY, V. L., ARNOLD, L. S., & PRUESSE, M. G. (1980). MMPI profiles of men referred for a pretrial psychiatric assessment as a function of offense type. *Journal of Clinical Psychology*, 36, 410–417.
- ROUNDTREE, G. A., & FAILY, A. C. (1980). The impact of educational programs on acts of aggression and rule violations in a female prison population. *Corrective & Social Psychiatry & Journal of Behavior Technology, Methods & Therapy*, 26, 144–145.
- SANTAMOUR, M. B., & WEST, B. (1982). The mentally retarded offender: Presentation of the facts and discussion of the issues. In M. B. Santamour & P. S. Watson (Eds.), *The retarded offender*. New York: Praeger.
- SCHERENBERGER, R. C. (1982). Comprehensive programming. In M. B. Santamour & P. S. Watson (Eds.), *The retarded offender*. New York: Praeger.
- SCHILT, J. (1979). The mentally retarded offender and criminal justice personnel. *Exceptional Children*, 46, 16–22.
- SHAPIRO, A. (1968). Delinquent and disturbed behavior within the field of mental deficiency. In A. U. S. de Reuck & R. Porter (Eds.), *The mentally abnormal offender*. London: J. & A. Churchill, Ltd.
- STEINER, J. S. (1984). Group counseling with retarded offenders. *Social Work*, 29, 181–185.
- SUTKER, P. B., & ALLAIN, A. N., JR. (1987). Cognitive abstraction, shifting, and control: Clinical sample comparisons of psychopaths and nonpsychopaths. *Journal of Abnormal Psychology*, 96, 73–75.
- TAMMANY, J. M., EVANS, R. G., & BARRETT, R. W. (1990). Personality and intellectual characteristics of adult male felons as a function of offense category. *Journal of Clinical Psychology*, 46, 906–911.
- TARJAN, G. T., WRIGHT, S. W., EYMAN, R. K., & KEERAN, C. V. (1973). Natural history of mental retardation: Some aspect of epidemiology. *American Journal of Mental Deficiency*, 77, 367–379.
- WILSON, J. Q., & HERSTEIN, R. (1985). *Crime and human nature*. New York: Simon & Schuster.

JAMES R. P. OGLOFF
DAVID G. SCHAEFER

CRONBACH, L. J. (1916–) Lee J. Cronbach was born in Fresno, California. He received his A. B. from Fresno State College in 1934, his M. A. from the University of California, Berkeley, in 1937, and his Ph.D. from the University of Chicago in 1940. He was a professor of education and psychology at the University of Illinois until 1963. At age 4 Cronbach was identified as one of Lewis TERMAN's intellectually gifted children and later, on moving to Stanford University as professor in 1963, became a principal coinvestigator on the continuing Terman longitudinal study (see TERMAN'S GIFTEDNESS STUDY).

Cronbach's career includes contributions to the improvement of inquiry in the social sciences and education. Although his principal line of scholarship has concerned measurement theory and methodology aimed at the development and evaluation of educational and psychological measures and assessment, his work spans several substantive domains of psychology,

educational research and evaluation, and the philosophy of social science. In their time, his books did much to define their fields of psychological testing, educational psychology, and evaluation methodology. His chief contributions in relation to intelligence theory, research, and measurement include the following: the extension of reliability theory, particularly Coefficient Alpha (usually called *Cronbach's alpha*) and the broad range of generalizability theory beyond it; the development of validity theory, particularly the introduction of the concept of construct validation; leadership in the creation of scientific and professional standards for tests; the application of utility theory to test use and the distinction between validation for classification versus selection; the combination of correlational and experimental methods and concepts in the invention of the aptitude-treatment interaction paradigm; the separation of between-class and within-class regression analysis of aptitude-learning relationships; the development of methodology for the measurement of intellectual development and change; the clarification of issues in public policy debates about mental tests; and the formulation of new models for social research, emphasizing rich description of situated phenomena rather than the hypothetico-deductive search for formal laws.

Cronbach's methodological contributions have had far-reaching effects on research and practice in intelligence and mental testing. His scholarly writings have not only established appropriate methods for test development, evaluation, and use but have also revolutionized many aspects of the field. His textbooks remain standard references on intelligence in education and the role of mental testing in the modern world.

BIBLIOGRAPHY

- CRONBACH, L. J. (1946). Response sets and test validity. *Educational and Psychological Measurement*, 6, 475-494.
- CRONBACH, L. J. (1951). Coefficient Alpha and the internal structure of tests. *Psychometrika*, 16, 297-334.
- CRONBACH, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- CRONBACH, L. J., & MEEHL, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281-303.

CRONBACH, L. J. (1991). Methodological studies—A personal retrospective. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 385-400). Hillsdale, NJ: Erlbaum.

SNOW, R. E., & WILEY, D. E. (1991). Straight thinking. In R. E. Snow & D. E. Wiley (Eds.), *Improving inquiry in social science: A volume in honor of Lee J. Cronbach* (pp. 1-11). Hillsdale, NJ: Erlbaum.

RICHARD E. SNOW

CROSS-CULTURAL VARIATIONS IN INTELLIGENCE No single definition of the concept of intelligence has been agreed upon. This diversity of views is compounded when variations due to different cultural perspectives are introduced. Numerous alternative culturally based meanings of the concept have been proposed by those working with other cultures. Some have even questioned the very existence of the concept of intelligence as it is commonly understood in Western psychological science.

In contrast, others have emphasized similarities rather than differences. It is asserted that all human beings as members of a single species must share some basic psychological processes; if we did not, then linguistic communication and cooperative, coordinated action across cultural boundaries would not be possible.

These contrasting sets of views provide the poles of a dimension along which material presented in this article will range. The article presents points of view that emphasize both differences and similarities in intelligence across cultures. And it takes seriously the role of culture as an important element in understanding the notion of intelligence in two ways: as a definer of the goals of human intellectual development (i.e., what it means to be an intelligent person in a particular culture) and as a shaper of development toward those valued goals (i.e., as a promoter or inhibitor of the individual's attainment of intelligence).

Despite these variations in views about the concept of intelligence across cultures, it is useful to adopt a working definition at the outset: *Intelligence* refers to that cluster of abilities developed by individuals in adaptation to their ecological and cultural contexts. This

working definition is silent with respect to whether intelligence is identical in all individuals and all cultural groups or differs (in quality and in quantity) across individuals and cultural groups.

This definition does represent one particular point of view: that the purpose of understanding and assessing intelligence is to describe the intelligence of a particular cultural group. At least one other, alternative point of view exists: that the purpose is to predict the intellectual, academic or occupational achievement of individuals of various cultural backgrounds when they live in a Western society. In this case, the culture of the Western society would be an appropriate cultural frame of reference, rather than the culture of the particular group. From the perspective of cross-cultural psychology, which is taken here, the second approach is based on a number of assumptions that are usually considered ethnocentric, and hence it is neither adopted nor elaborated here.

The article will proceed with a brief outline of three theoretical positions regarding the understanding of human psychological diversity in general (absolutism, relativism, and universalism) and then apply them to the concept of intelligence specifically. It will then review the three main domains of research on intelligence across cultural groups: the cultural, the minority, and the ethnic approaches. Finally, it will examine issues of assessment and applications in the light of these distinctions and initial discussions.

APPROACHES TO UNDERSTANDING PSYCHOLOGICAL DIVERSITY

Most psychological phenomena can be studied in relation to culture from a variety of perspectives. Three such positions have been discussed in research on human diversity: absolutism, relativism, and universalism (Berry et al., 1992). The "absolutist" position is one that assumes that psychological phenomena are basically the same (qualitatively) in all cultures: "honesty" is "honesty," and "depression" is "depression," no matter where one observes it. From the absolutist perspective, culture is thought to play little or no role in either the meaning or display of psychological characteristics. Assessments of such characteristics are made using standard instruments (perhaps with lin-

guistic translation of test items); comparisons of scores obtained are made directly; and interpretations are made easily, without alternative culturally based views being taken into account.

In sharp contrast, the "relativist" approach is rooted in anthropology and assumes that all human behavior is culturally patterned; it seeks to avoid ethnocentrism by trying to understand people "in their own terms." Explanations of psychological variation are sought in the cultural context in which people have developed, and usually little interest is shown in the possible role of biology. Assessments are typically carried out employing the values and meanings a cultural group gives to a phenomenon and only rarely in more than one culture. Comparisons are judged to be problematic and ethnocentric, and are thus virtually never made.

A third perspective, one that lies somewhere between the two other positions, is that of "universalism." Here it is assumed that basic psychological processes are common to all members of the species (i.e., constituting a set of biological givens) and that culture influences the development and display of the behaviors that are rooted in these processes (i.e., culture plays different variations on these underlying themes). Assessments of psychological characteristics are based on the presumed underlying process, but test items are developed in culturally meaningful versions. Comparisons are made cautiously, employing a wide variety of methodological principles and safeguards (Lonner & Berry, 1986), and interpretations of similarities and differences are attempted that take alternative culturally based meanings into account.

APPROACHES TO UNDERSTANDING INTELLIGENCE ACROSS CULTURES

When the three general theoretical approaches described above are used to examine the literature on intelligence across cultures, a reasonably good match is found between these perspectives and the actual research carried out to date.

Absolutism. Most of the early work on intelligence and its assessment across cultures was typically carried out from the absolutist perspective (Cole & Scribner, 1975; Irvine & Berry, 1988; Vernon, 1969).

Typically a common standard test was employed with samples from a number of cultural groups, with very little account being taken of linguistic or other cultural variations. At most, linguistic translation of instructions and test items was provided. For example, the Binet and Wechsler tests for intelligence quotient (IQ) have been employed in numerous countries around the world, sometimes translated, sometimes not (Kamin, 1983). Test content (both the subscales and items within subscales) was rarely altered for the population tested. Less wide-ranging tests (e.g., RAVEN PROGRESSIVE MATRICES and Porteus Mazes) have also been used in this way (e.g., Irvine, 1959). Direct comparisons of scores, leading to assertion of higher or lower intelligence in one group in relation to others, were not uncommon. This approach has been taken both internationally (across nation-states) and domestically (across ethnocultural groups within nation-states). The contemporary literature demonstrates the continuing use of this approach, as in American–Japanese comparisons (e.g., Lynn, 1982) and in African-American–Asian-American–American white comparisons (e.g., Jensen & Reynolds, 1982; Vernon, 1982). A similar approach has been taken using measures that are considered by the researchers to assess some fundamental aspect or correlate of intelligence, perhaps rooted in human biology. For example, H. J. EYSENCK (1988) has used the “average evoked potential” (brain waves) and A. R. JENSEN (1988) has used “speed of information processing” (REACTION TIME) as measures that they believe may avoid some of the problems of comparison across cultural groups.

Relativism. The anthropological tradition of seeing people in their own terms has led some psychologists to seek out the indigenous meaning of intelligence prior to attempting assessment of their intelligence (Berry, Irvine, & Hunt, 1987). This position has also been advocated on social-policy grounds, in order to avoid harming individuals and groups through mismeasurement and misinterpretation of their intelligence (e.g., Berry, 1972). Whatever the reason for adopting the relativist position, a first step is to try to understand what is meant by intelligence within the local cultural context. Studies in Africa (e.g., Dasen, 1984; Serpell, 1989; Wober, 1974) and the Arctic (Berry & Bennett, 1992) have revealed meanings that are notably different from the “fast, an-

alytic, and purely cognitive” cluster of abilities that predominate in Western academic psychology. Rather, elements of thoughtfulness, caution, respect, and social understanding frequently characterize the views in many other cultures (Berry, 1984), as well as in daily life in Western cultures (Wassmann & Dasen, 1993). Within the relativist research tradition, given the divergent meanings of what it is to be intelligent, comparisons of the level of intelligence developed by individuals become impossible, at least for quantitative comparisons. At most, qualitative comparisons can be made in order to document the extent of variations in the meanings of intelligence as a basis for avoiding simple quantitative comparison. If individuals who are raised in cultures with divergent views about intelligence and who are routinely nurtured toward those valued goals are directly compared on a common standard test, then the conclusions will be not only invalid but also harmful. This relativist view is considered by many to be applicable equally to studies across cultures and studies of ethnocultural groups within plural societies (Berry, 1993).

Universalism. Drawing to some extent on the contrasting perspectives of absolutism and relativism, the universalist approach attempts to establish a valid basis for comparison (rooted in the assumption of specieswide shared psychological processes associated with the absolutist position), while respecting local cultural variations in meanings and competence (rooted in the relativist position). An important early insight of this position was articulated by G. Ferguson (1956, p. 121): “Cultural factors prescribe what shall be learned and at what age: Consequently, different cultural environments lead to the development of different patterns of ability.” This view was elaborated by S. H. Irvine and J. W. Berry (1988) into a “law of cultural differentiation,” which is based on the ecocultural approach (Berry, 1976) to understanding variation across cultures in human abilities. This approach hypothesizes what pattern of abilities is likely to be useful in adapting to a particular ecological setting, using the notion of “ecological demands” (i.e., specifying those abilities that need to be developed in order to function), and by searching for the set of “cultural aids” that will nurture the development of those abilities (Berry, 1966). From this perspective, the initial range of possible abilities is panhuman in nature, but

which abilities become developed and to what extent are under the control of local ecological and cultural circumstances and practices.

Within this perspective, variations in spatial abilities and cognitive styles (Berry, 1976) and in the attainment of Piagetian stages (Dasen, 1972) have been successfully predicted across large variations in ecological and cultural contexts, such as hunter-gatherers and horticulturalists. The use of this perspective is considered to be equally applicable to understanding the intelligence of ethnocultural groups living in plural societies (Berry, 1993).

Comparisons of intelligence are difficult to accomplish within the universalist framework, but some principles and procedures have been developed that permit general comparisons (Irvine, 1986; Irvine & Carroll, 1980). Indications of item difficulty, item correlations, and the factorial structure of the test scores are the usual bases of comparison prior to, and sometimes instead of, comparing the mean scores obtained by the different cultural groups (Poortinga & Van der Flier, 1988).

APPROACHES TO THE ASSESSMENT OF INTELLIGENCE

The recognition of the role of cultural factors in the definition and development of intelligence has led, even within the absolutist perspective, to some consideration of how to make comparisons of intelligence more "fair" or "unbiased." Initially, the search focused on "culture fair" tests of intelligence (e.g., Cattell & Cattell, 1963), but it soon became clear that such a search was futile because culture and intelligence are intertwined in so many ways (see CULTURE-FAIR AND CULTURE-FREE TESTS).

First, the language of test administration (both the language of instruction and the language of the test content) is usually English or some other European language. Individuals whose mother tongues are not those languages or who speak a nonstandard dialect of those languages would not be presented with the same task to do, for they would first have to try to figure out what they are supposed to do and then figure out what the test item means. Linguistic translation is thus a minimal prerequisite for "fairness," but it is far from sufficient (Brislin, 1986). Content translation is thus

also required. For example, an IQ test item that assesses one's knowledge of the distance between two cities presupposes that one knows what a city is (which is not always the case for rural people); knows that distance is to be estimated by space (rather than time taken to travel); knows that it is direct distance, not by road or foot (for which pathways may meander around mountains or rivers); knows the particular unit of space to be used (kilometers, land miles, nautical miles, leagues, etc.); and knows the names of the particular cities mentioned in the test term. Each of these aspects is linked to particular cultural assumptions, which may not be shared by the test maker, the test administrator, and the test taker. Obviously, the content of the item has to be changed if it is to be a fair assessment. But even this is not enough, because for many people in the world cities are unimportant, and travel to or between cities is out of the question. In their cases, of what possible relevance is such a question to their intelligence?

To provide another route to fairness, attempts have been made to replace verbal item content (and test instructions) by nonverbal materials. This has been most evident in the introduction of pictorial or figural test items (as in the Cattell test and in the Raven Progressive Matrices). Even so, the understanding of such materials requires a form of literacy and acceptance that real-life objects and forms can be represented by lines on a two-dimensional paper surface. Both literacy and pictorial representation are cultural products that are learned in some cultures but not in others; moreover, styles of pictorial representation are known to vary greatly across cultures (Deregowski, 1980).

Beyond linguistic and content translation to make tests more fair, there is (as implied earlier) the need to match the assessment instrument to both the cultural meaning and the indigenous nurturing of the development of intelligence. If a certain set of abilities is valued and nurtured in a particular cultural group, and an intelligence test assesses a different set of abilities (based on what is valued and nurtured in the culture in which the test arose), then the degree of mismatch will be the degree of bias or unfairness of the test for that particular cultural group.

In sum, the search for culture-fair tests, by attempting to make them "culture-free" or "culture-reduced," is extremely difficult; both linguistic and

content translation are time-consuming and may not result in the equivalent forms of the test that are necessary to make valid comparisons. The view that the only really fair test is one that assesses intelligence in a way that matches the meaning and development of intelligence in a particular cultural group makes the search not only difficult but possibly futile as well.

DOMAINS OF ASSESSMENT OF INTELLIGENCE

Two domains of work have been mentioned in this article: international comparisons across cultural groups and domestic comparisons across ethnocultural groups living together within pluralistic societies. In principle, it is possible to take into account all three theoretical perspectives in both of these domains. Yet, in practice, the tendency has been for work in the latter (domestic) domain to ignore or seriously undervalue the cultural meanings and contexts of ethnocultural groups. This practice may be termed the "minority" approach to understanding intelligence in ethnocultural groups, and it is generally associated with the absolutist perspective (Samuda, 1975). While initially the absolutist perspective was dominant in international studies of intelligence, the trend toward using relativist and universalist perspectives is increasing. This may be termed the "cultural" approach. Given the increasing sensitivity to, and awareness of, the human consequences of misunderstanding intelligence across cultures, it is likely that assessment of ethnocultural groups within pluralistic societies will shift away from the minority approach and adopt one based on the cultural perspective; this has been termed the "ethnic" approach (Berry, 1985).

The minority approach to understanding ethnocultural groups has tended to take little account of their cultural circumstances (see Vernon, Jackson, & Messick, 1988, for numerous examples of ignoring such circumstances). Most psychological research and assessments hardly mention the ecological settings or cultural qualities that provide the contexts for the development of intelligence (or of any other psychological quality). If such an account were taken, many ethnocultural groups (e.g., blacks in the United States, Vietnamese in Canada, aborigines in Australia, Turks in Germany, North Africans in France, and Asians in

Great Britain) would be seen in their own terms and their intelligence (and other) test performance would be seen as an adaptive outcome of their ecological and cultural circumstances rather than as some deviation from a "mainstream" society.

A variety of "deviation" concepts have been used by those taking the minority approach. These have been described and criticized by D. McShane and J. W. Berry (1988). These "d-models" all propose explanations of variations in intelligence based on some "deviation" of the "minority" from the "mainstream." These include "genetic deficit" (postulating some genetic deficit for the deviation), "physiological defect" (such as the effects of alcohol or the presence of auditory or visual problems), "disadvantage" (the presence of poverty, poor nutrition, and poor health care), "deprivation" (a combination of psychological deficit and disadvantage), "cultural disorganization" (the marginalization or exclusion of minorities from full participation in the mainstream), and "disruption" (such as the uprooting of children from their parents, as when they are sent to boarding schools or hospitalized for a prolonged period).

The hallmark of all these models is that minorities are seen in terms of the mainstream and judged (usually negatively) in those terms. The alternative is a "difference model," in which psychological diversity is understood and interpreted in terms of the life circumstances (both ecological and cultural) of the group. This model requires the reconceptualization of such groups from "minority" into "ethnocultural" or "ethnic" groups (Berry, 1985, 1993). From this perspective, groups living together in pluralistic societies are viewed as legitimate forms of social and cultural life, and their cognitive abilities are assessed and interpreted in terms of their own cultures. This transformation can take place only when the cultural approach replaces the minority approach and emerges as the ethnic approach to psychological assessment of groups in pluralistic societies.

CONCLUSION

This article has attempted to portray variations in how psychologists have approached the understanding and assessment of intelligence across cultures and to evaluate these various approaches. The evidence sug-

gests that the concept of intelligence is intimately linked to the concept of culture in a variety of ways. First, different cultural groups are known to value and nurture different aspects of intelligence, leading to different indigenous conceptions of intelligence and to different outcomes with respect to the intelligence that is developed. Second, psychologists (mainly working within Western cultures and espousing a Western cultural view of intelligence) have attempted to assess intelligence in other cultural groups with tests that can be seen as invalid, or biased, as instruments for gauging the intelligence of other peoples. A likely solution to these problems is not to seek culture-free or culture-fair concepts of measures of intelligence but to develop and work with culture-linked concepts and measures. These would match indigenous views and competencies and hence be more valid and less biased. This alternative approach was termed "cultural" when taken internationally across cultures and "ethnic" when employed with groups living together in pluralistic societies. In this latter case, a reconceptualization is needed of such groups as "ethnocultural" (with functioning cultures of their own), rather than as "minorities" (usually seen as deviations from some "mainstream" culture).

(See also: CULTURE AND INTELLIGENCE; EASTERN VIEWS OF INTELLIGENCE.)

BIBLIOGRAPHY

- BERRY, J. W. (1966). Temne and Eskimo perceptual skills. *International Journal of Psychology*, 1, 207-229.
- BERRY, J. W. (1972). Radical cultural relativism and the concept of intelligence. In L. J. Cronbach & P. J. D. Drenth (Eds.), *Intelligence and cultural adaptation*. The Hague: Mouton.
- BERRY, J. W. (1976). *Human ecology and cognitive style: Comparative studies of cultural and psychological adaptation*. London: Sage Publications.
- BERRY, J. W. (1984). Toward a universal psychology of cognitive competence. *International Journal of Psychology*, 19, 335-361.
- BERRY, J. W. (1985). Cultural psychology and ethnic psychology: A comparative analysis. In I. Reyes-Lagunes & Y. Poortinga (Eds.), *From a different perspective*. Lisse, Netherlands: Swets & Zeitlinger.
- BERRY, J. W. (1993). An ecocultural perspective on human diversity. In E. Trickett (Ed.), *Human diversity*. San Francisco: Jossey-Bass.
- BERRY, J. W., & BENNETT, J. A. (1992). Cree conceptions of cognitive competence. *International Journal of Psychology*, 27, 73-88.
- BERRY, J. W., IRVINE, S. H., & HUNT, E. B. (EDS.). (1987). *Indigenous cognition: Functioning in cultural context*. Dordrecht, Netherlands: Nijhoff.
- BERRY, J. W., POORTINGA, Y., SEGALL, M., & DASEN, P. (1992). *Cross-cultural psychology: Research and applications*. New York: Cambridge University Press.
- BIESHEUVEL, S. (1943). *African intelligence*. Johannesburg: South African Institute of Race Relations.
- BRISLIN, R. (1986). The wording and translation of research instruments. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research*. London: Sage Publications.
- CATTELL, R. B., & CATTELL, A. K. (1963). *Test of g: Culture Fair, Scale 3*. Champaign, IL: Institute for Personality and Ability Testing.
- COLE, M., & SCRIBNER, S. (1975). *Culture and thought*. New York: Wiley.
- DASEN, P. R. (1972). Cross-cultural Piagetian research: A summary. *Journal of Cross-cultural Psychology*, 7, 75-85.
- DASEN, P. (1984). The cross-cultural study of intelligence. Praget and the Baoulé. *International Journal of Psychology*, 19, 407-434.
- DEREGOWSKI, J. (1980). Illusions, patterns, and pictures: A cross-cultural perspective. London: Academic Press.
- EYSENCK, H. J. (1988). The biological bases of intelligence. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context*. New York: Cambridge University Press.
- FERGUSON, G. (1956). On transfer and the abilities of man. *Canadian Journal of Psychology*, 10, 121-131.
- IRVINE, S. H. (1959). Figural tests of reasoning in Africa: Studies in the use of Raven's Progressive Matrices across cultures. *International Journal of Psychology*, 4, 217-228.
- IRVINE, S. H. (1986). Cross-cultural assessment: From practice to theory. In W. J. Lonner & J. W. Berry (Eds.), *Field methods in cross-cultural research*. London: Sage Publications.
- IRVINE, S. H., & BERRY, J. W. (1988). The abilities of mankind: A reevaluation. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context*. New York: Cambridge University Press.
- IRVINE, S. H., & CARROLL, W. K. (1980). Testing and assess-

- ment across cultures. In H. C. Triandis & J. W. Berry (Eds.), *Handbook of cross-cultural psychology: Vol. 2. Methodology*. Boston: Allyn & Bacon.
- JENSEN, A. R. (1988). Speed of information processing and population differences. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context*. New York: Cambridge University Press.
- JENSEN, A. R., & REYNOLDS, C. R. (1982). Race, social class, and ability patterns on the WISC-R. *Personality and Individual Differences*, 3, 423-438.
- KAMIN, L. J. (1983). *The science and politics of IQ*. Harmondsworth, England: Penguin.
- LONNER, W. J., & BERRY, J. W. (EDS.). (1986). *Field methods in cross-cultural research*. London: Sage Publications.
- LYNN, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature*, 297, 222-223.
- MC SHANE, D., & BERRY, J. W. (1988). Native North Americans: Indian and Inuit abilities. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context*. New York: Cambridge University Press.
- POORTINGA, Y., & VAN DER FLIER, H. (1988). The meaning of item bias in ability tests. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context*. New York: Cambridge University Press.
- SAMUDA, R. (1975). *Psychological testing of American minorities*. New York: Harper & Row.
- SERPELL, R. (1989). Dimensions endogènes de l'intelligence chez les A-chewa, et les autres peuples africains. In J. Ratschitzky, M. Bossel-Lagos, & P. Dasen (Eds.), *La Recherche interculturelle*. Paris: l'Harmattan.
- VERNON, P. E. (1969). *Intelligence and cultural environment*. London: Methuen.
- VERNON, P. E. (1982). *The abilities and achievements of orientals in North America*. New York: Academic Press.
- VERNON, P. E., JACKSON, D. N., & MESSICK, S. (1988). Cultural influences on patterns of abilities in North America. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context*. New York: Cambridge University Press.
- WASSMANN, J., & DASEN, P. (EDS.). (1993). *Alltagswissen/Les savoirs quotidiens/Everyday cognition*. Freiburg: Universitätsverlag Freiburg.
- WOBER, M. (1974). Towards an understanding of the Kiganda concept of intelligence. In J. W. Berry & P. Dasen (Eds.), *Culture and cognition*. London: Methuen.

JOHN W. BERRY

CULTURE-FAIR AND CULTURE-FREE TESTS

Psychological tests are samples of what people know and know how to do. Tests are *not* samples of "innate" intelligence or culture-free knowledge (an oxymoron). In other words, intelligence and ability tests are samples of human cultural knowledge, acquired across development. There is no such thing as a culture-free test.

Tests vary in the degree to which the knowledge and skills they sample are specific to a particular culture or more general to many or most cultures (Grubb & Ollendick, 1986). The specificity of items on a test can vary from skills that every normal member of the human species can be expected to acquire at some point in development (e.g., naming primary colors) to knowledge that only a small number of educated people in a specific culture is likely to know (e.g., defining the word "obfuscation").

Because the major purpose of intelligence and ability tests is to predict future achievements in school and occupations in a particular culture (e.g., the United States), the cultural content of test items is quite appropriate. The learning of age-appropriate vocabulary items and arithmetic operations predicts later verbal and mathematical achievements, even though the specific vocabulary and arithmetic items are culturally defined. School and job success are the criteria of achievement that tests are designed to predict; the fact that the criterion and the predictor share the same culture makes the culture loading of tests not only acceptable but often preferable to tests with less specific cultural content.

It is inappropriate to interpret test scores, however, as measures of "innate" ability or ability to perform in a culture other than the one in which the person was reared. One has only to imagine being tested in another language about the history and culture of another nation.

All tests assume that people taking the tests have had adequate opportunities to acquire the skills and knowledge tested (Vernon, 1979). That is, it is inappropriate to test someone who has had little or no exposure to the domain of material of which the test material is a sample (Jensen, 1980). As an extreme example, it makes no sense to test a blind person's mastery of visual tasks or a profoundly deaf person's oral

responses to spoken vocabulary. As a less extreme example, testing a native German speaker in English makes sense only if the person will have to perform in school or on the job in English. In these cases, it is easy to see that one would not conclude anything about “innate” intelligence based on tests of skills and knowledge to which the person had little or no exposure. One would have an accurate assessment, however, of how well each of these persons would be likely to perform tasks, such as these in criterion school or job settings, that share the same culture as the tests.

CULTURE-REDUCED TESTS

Something of a misnomer, *culture-reduced* tests require knowledge and skills that are common to more than one or to many cultures. They are not so much culture-reduced as they are general to more than one culture. Finding the correct piece that fits into a puzzle space, recognizing familiar items in a memory test, and matching similar stimuli in an array of diverse objects are examples of skills that are common to many cultures. Such tests are less culturally specific than vocabulary, arithmetic, and information subtests commonly found on the most widely used intelligence tests.

Among the best known culture-reduced tests are the RAVEN PROGRESSIVE MATRICES, a test of abstract reasoning. Examples of matrix items are shown in Figure 1.

Matrix items are puzzles with missing pieces or arrangements of figures with sequential changes. The problem is to figure out which piece fits the blank space or what the next change in the series of figures should be. Although the Raven test seems to require little or no language, it is closely related to vocabulary and other tests of general cultural knowledge (Jensen, 1980). Although the matrices seem to require spatial skills, performance on this test of abstract reasoning is less related to spatial than to verbal ability.

Contrary to most people’s intuition, ethnic minority children in the United States have been shown repeatedly to score better on culturally loaded material than on more abstract tests that seem to sample less directly from the dominant culture (Jensen & McGurk, 1987). For abstract items, such as those on the Raven Progressive Matrices (see Figure 1), understanding the task is more than half the battle. In one study, many of the ethnic minority children in the sample did not understand the instructions to the “game,” and thus could not solve the problem (Scarr, 1981). Whereas instructions for culturally loaded items are clear from everyday experience (“Why do we have ovens?”), instructions for culturally reduced items are not so obvious (“Choose the figure that completes the sequence.”). In a large study of African-American twins, test instructions were altered, making them exceedingly redundant. In this circumstance, African-American adolescents scored higher on culturally reduced

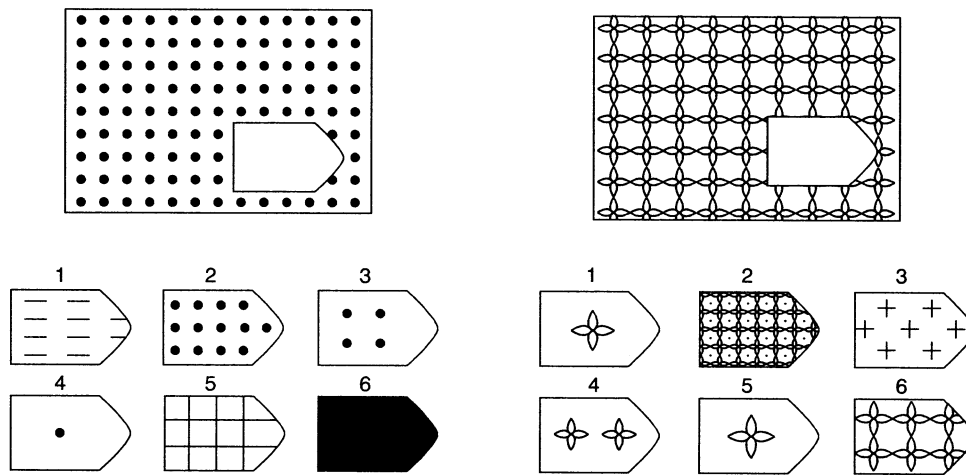


Figure 1
Raven Progressive Matrices items

than culturally loaded items, and the larger group differences were on culturally loaded tests rather than on culturally reduced ones (Scarr, 1981). Of course, doing well in school requires listening to and following instructions, so it is not surprising that poor test performance predicts lower school achievement.

CAN A TEST EVER BE CULTURALLY FAIR?

Cultural psychologists have questioned the possibility of culture-fair tests of intelligence for people to whom the idea of an artificial test situation is very strange. People who have never been to school do not seem to grasp the game-like testing situation in which, for no apparently good reason, they are supposed to bring their best intellectual skills to play (Stigler, Schweder, & Hurt, 1990). Jacqueline Goodnow (1990) has further questioned the assumptions underlying Western intelligence tests, assumptions that we share implicitly; for example, that fast is better than slow, that simple reasoning is better than more complex or indirect reasoning, that every problem has a right answer. Even nonverbal, culturally reduced tests are based on this set of assumptions which may not be shared by culturally different peoples. In this sense, even culturally reduced tests cannot be culturally fair.

Culturally reduced tests have another problem: They do not correlate as highly with criterion performances in school or on jobs as do culturally loaded tests, because they sample a narrow range of reasoning, rather than a broad domain of culturally relevant knowledge and skills. Although tests such as the Raven Matrices may seem more fair because they sample skills that are learned by nearly everyone (or no one), puzzle-like tests turn out to have their own limitations.

CULTURE LOADING VERSUS CULTURE BIAS

Tests that are culturally loaded are not necessarily culturally biased. In psychometric (mental measurement) terms, culture bias occurs when one group's achievements are more poorly predicted by the tests than another group's achievements (Drasgow, 1987). In general, intelligence and ability tests are not cultur-

ally biased by testing criteria; that is, they predict school and job performance just as well in all groups tested (Boehm, 1977). Although some ethnically different and lower socioeconomic groups score, on average, lower than other groups on intelligence and ability tests, the tests accurately predict their lower performance on criteria of schools and jobs. Because tests predict as accurately for all groups, they are not considered biased against any group.

Tests do not tell us *why* some groups score lower than others. They tell us only that people who score lower on tests are likely to score lower on the criteria as well, regardless of the group they come from. Research on ethnic-minority children shows that the most likely explanation for their lower average performance is that they have less exposure than majority-group children to "the culture of the tests and the culture of the schools" (Scarr & Weinberg, 1983; Weinberg, Scarr, & Waldman, 1992). Because tests are loaded with knowledge acquired in a particular culture, lack of opportunity to learn that culture necessarily reduces performance on the test and predicts poorer performance on other criteria that depend on that cultural knowledge.

The average difference in test performance between ethnic groups is not due to *internal biases* in the items within the tests. Critics of testing, who rely on intuitive appraisals of the culture loading of test items (e.g., "define minotaur"), have simply not examined the overwhelming evidence that the difficulty levels of items on intelligence tests are not a function of their subjective cultural loadings. Items that are difficult in one group are approximately as difficult in another group.

WHAT IS TEST BIAS?

Some groups do not score, on average, as highly on ability and aptitude tests as other groups. The lack of equality in outcome of testing often leads critics to infer test bias, but this is incorrect. Bias refers to the systematic under- or overestimation of a population parameter by a statistic based on samples drawn from that population (Jensen, 1980, p. 375). Psychometric bias refers to systematic errors in the *predictive* or *construct* validity of test scores in one or more groups, such as racial, social class, or gender groups. Predictive

validity of a test can be biased if the slopes of regression lines between the predictor test (e.g., IQ) and the criterion (e.g., school achievement score) differ for two groups. Bias in construct validity occurs when the relationships between test scores and a theoretically related construct differ in two or more populations. Construct validity has two aspects: external and internal validity. External biases, such as predictive validity, can be seen in different slopes in the regression of conceptually related constructs (such as speed in learning a trade) on intelligence test scores in two populations. Internal biases refer to lack of content validity or errors in sampling the knowledge or skills to be tested. Thus, there are several ways in which tests themselves can be biased. Any test can be administered in a biased manner or administered to an inappropriate group (e.g., speakers of a different language). Jensen (1980) also includes differences in test-score intercepts and in standard errors between two groups as evidence of test bias in predictive validity.

IQ TESTS

A vast research literature developed in the 1970s and 1980s to test the alleged bias in IQ, aptitude, and school achievement tests. There is overwhelming evidence that, by psychometric criteria, IQ tests are not biased against African Americans, Hispanics, Asians, or other ethnic groups. Predictive validity and construct validity coefficients and regression slopes are similar in African-American, Hispanic, Asian-American, and Euro-American groups, despite average group differences. Only on criteria of content validity, where judgments about appropriateness of content are made, can the test be challenged on psychometric grounds, but even here, items judged a priori to be culturally biased are usually found not to be so.

MISUNDERSTANDINGS OF BIAS IN TESTING

Tests do not guarantee that we will score above average, like the children of Lake Wobegone. A mean difference between the IQ test scores of two groups is not *prima facie* evidence of test bias. The criterion (e.g., school grades) may show the same average difference as the test between the two groups and parallel regres-

sion slopes, as shown in Figure 2. Even the fact that a test has been standardized primarily on one population and administered to another may or may not introduce bias. The cultural loading of a test is not itself evidence of bias, especially when the criteria to which the test predicts share the same cultural loading.

Eyeballing the items in an IQ test is not an adequate means to detect cultural bias. In the Chicago PACE case that tried and failed to exclude IQ test from schools, Judge Grady selected eight items from the Wechsler Intelligence Test for Children and one item from the Stanford-Binet Test of Intelligence as culturally biased against black children, based entirely on his personal assessment (Elliott, 1987). Later research showed that these items were no more difficult for African-American children than other items on the test, and that the race differences on these items were smaller than those on many other items with less obvious cultural content. Several investigators have, in fact, shown that more culturally loaded items are often easier than more abstract ones for both black and white children.

UNFAIR USES OF TESTS

Judgments about unfair uses of tests (Williams & Mitchell, 1991) in selection for educational and occupational purposes depend not on test bias but on outcomes of test use that are considered, by some philosophical criterion, to be undesirable (Hunter & Schmidt, 1976, 1978). Tests that meet all psychometric criteria for being unbiased often result in unequal outcomes for different groups in a selection situation (Hunter, Schmidt, & Huneft, 1979). Results of biased and unbiased tests can be adjusted to yield equal selection outcomes for groups with different average test scores, if equal outcomes are the goal. But biased tests are not necessarily unfair, nor are unbiased tests necessarily used in fair ways (Jensen, 1980).

Mercer (1979) proposed that IQ test scores of children from ethnic minority groups and socially disadvantaged children be adjusted to take into account their lesser exposure to the dominant culture and fewer opportunities to learn the material that appears on tests. Thus, an African-American child from a low-income family in an inner-city neighborhood whose IQ score is 89 has a "learning potential" above that score.

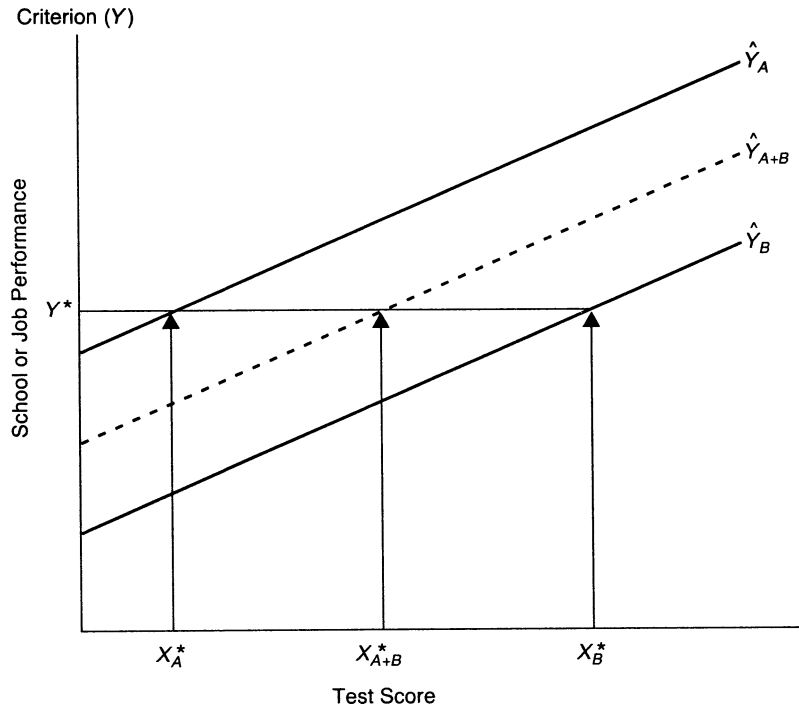


Figure 2
A biased test in which the major and minor groups, A and B, have regression lines differing only in intercepts. The slopes of the regressions are the same.

Mercer developed a System of Multicultural Pluralistic Assessment (SOMPA) to adjust the scores of disadvantaged children on standard IQ tests. The more disadvantaged a child's family and cultural background, the greater the upward adjustment of the IQ score. In principle, this seems to be a fair way to assess intelligence for children who have not had good opportunities to learn the culture of the tests.

One problem with SOMPA is that the adjusted scores do not predict school performance: A child who scores IQ 89 performs in reading and mathematics like other children with IQ scores in that range, regardless of the reasons for achieving that score. It could be that a disadvantaged child has more potential for intellectual change than an advantaged child, if the disadvantageous environment is changed dramatically, but the test score is a good prediction unless there is dramatic change. Another problem is that using adjusted IQ scores does not significantly reduce the proportion of minority children assigned to special education classes (Heflinger, Cook, & Thackrey, 1987).

Other attempts to alter test content by restricting samples of knowledge and skills to those that show the fewest ethnic differences, such as the K-ABC test, result in tests that are heavily loaded on rote memory rather than more general intelligence (Jensen, 1984; Witworth & Gibbons, 1986).

REACTION-TIME TESTS

Ever since Francis GALTON tried (and failed) to construct an intelligence test from simple sensory and perceptual responses, scientists have sought the ultimate culture-fair test—one that measures brain responses rather than acquired knowledge. Hans Eysenck and his colleagues (Frearson & Eysenck, 1987) and Arthur Jensen (1993) have reported moderate correlations (.30–.50) between certain reaction time measures and general intelligence measures. In a simple task (no one gets wrong answers), faster reaction times are not related to IQ scores, but greater VARIABILITY in reaction times is correlated with lower IQ scores. Both Eysenck and

Jensen interpret this result to mean that variable reaction times indicate unreliability in the nervous system, and that an unreliable nervous system also leads to lower IQ scores. Further developments may show that choice reaction time measures offer a less culturally loaded measure of intelligence, if the validity of the measure can be improved above the currently moderate levels.

DYNAMIC ASSESSMENT

Another method of avoiding cultural bias is the dynamic assessment of learning potential (Gupta & Coxhead, 1988). Dynamic assessment uses a novel learning task to avoid the confounding influence of previous learning on present performance. Dynamic assessment tests how well a child learns a task or a concept to which the child has never been exposed. The assessment techniques have been developed especially for children who are suspected of mental retardation, and those from ethnic minority groups. The future of dynamic assessment in the real world of schools and jobs is not clear, as the procedures are time-consuming and lack extensive evidence for validity in predicting important external criteria (Wurtz, Sewell, & Manni, 1985).

Culture-fair tests, whether matrices or learning tasks, have a role in the assessment of intelligence in some situations and for some populations. It is unlikely that they will replace culturally loaded tests that predict school and job performance better. It is critical to recall, however, that intelligence tests of any kind sample knowledge and skills that have been acquired, either directly as in culturally loaded tests, or indirectly as in learning tasks and matrix problems. The only real departure from standard assessment techniques is the use of variability in reaction-time tasks. Whether these measures will prove useful in predicting criteria of schools and jobs remains to be seen.

(See also: GROUP TESTS.)

BIBLIOGRAPHY

- BOEHM, V. R. (1977). Differential prediction: A methodological artifact? *Journal of Applied Psychology*, 62, 146–154.
- DRASGOW, F. (1987). Study of measurement bias of two standardized psychological tests. *Journal of Applied Psychology*, 72, 19–29.
- ELLIOTT, R. (1987). *Litigating intelligence: IQ tests, special education, and social science in the courtroom*. Dover, MA: Auburn House.
- FREARSON, W. M., & EYSENCK, H. J. (1986). Intelligence and reaction time (RT) and a new “odd-man-out” RT paradigm. *Personality and Individual Differences*, 7, 808–817.
- GONZALES, E. (1982). A cross-cultural comparison of the developmental items of five ethnic groups in the Southwest. *Journal of Personality Assessment*, 46, 26–31.
- GOODNOW, J. (1990). Cultural assumptions about testing. In J. Stigler, R. Schweder, & R. Hurt (Eds.), *Cultural Psychology*. New York: Cambridge University Press.
- GRUBB, H. J., & OLLENDICK, T. H. (1986). Cultural-distance perspective: An exploratory analysis of its effect on learning and intelligence. *International Journal of Intercultural Relations*, 10, 399–414.
- GUPTA, R. M., & COXHEAD, P. (EDS.). (1988). *Cultural diversity and learning efficiency: Recent developments in assessment*. Hong Kong: St. Martin's Press.
- HALPIN, G., SIMPSON, R. G., & SHEILA, L. (1990). An investigation of racial bias in the Peabody Picture Vocabulary Test—Revised. *Educational and Psychological Measurement*, 50, 183–189.
- HEFLINGER, C. A., COOK, V. J., & THACKREY, M. (1987). Identification of mental retardation by the System of Multicultural Pluralistic Assessment: Nondiscriminatory or nonexistent? *Journal of School Psychology*, 25, 177–183.
- HICKMAN, J. A., & REYNOLDS, C. R. (1987). Are race differences in mental test scores an artifact of psychometric methods? A test of Harrington's experimental model. *Journal of Special Education*, 20, 409–430.
- HUNTER, J. E., & SCHMIDT, F. L. (1976). A critical analysis of the statistical and ethical implications of various definitions of test bias. *Psychological Bulletin*, 83, 1053–1071.
- HUNTER, J. E., & SCHMIDT, F. L. (1978). Differential and single group validity of employment tests by race: A critical analysis of three recent studies. *Journal of Applied Psychology*, 63, 1–11.
- HUNTER, J. E., SCHMIDT, F. L., FRANK, L., & HUNTER, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin*, 86, 721–735.

- JENSEN, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- JENSEN, A. R. (1984). The black-white difference on the K-ABC: Implications for future tests. *Journal of Special Education, 18*, 377-408.
- JENSEN, A. R. (1993). Spearman's hypothesis tested with chronometric information-processing tasks. *Intelligence, 17*, 47-77.
- JENSEN, A. R., & MCGURK, F. C. (1987). Black-white bias in "cultural" and "noncultural" test items. *Personality and Individual Differences, 8*, 59-68.
- KOH, T. H., ABBATIello, A., & MCLOUGHlin, C. S. (1984). Cultural bias in WISC subtest items: A response to Judge Grady's suggestion in relation to the PASE case. *School Psychology Review, 13*, 89-94.
- MCLOUGHlin, C. S., & KOH, T. H. (1982). Testing intelligence: A decision suitable for the psychologist? *Bulletin of the British Psychological Society, 21*, 49-55.
- MERCER, J. (1979). *System of multicultural pluralistic assessment technical manual*. Berkeley: University of California Press.
- MIELE, F. (1979). Cultural bias in the WISC. *Intelligence, 3*, 149-164.
- NAGLIERI, J. A., & JENSEN, A. R. (1987). Comparison of black-white differences on the WISC-R and the K-ABC: Spearman's hypothesis. *Intelligence, 11*, 21-43.
- OAKLAND, T. (1983). Concurrent and predictive validity estimates for the WISC-R IQs and ELPs by racial-ethnic and SES groups. *School Psychology Review, 12*, 57-61.
- POTEAT, M. G., WUENSCH, K. L., & GREGG, N. B. (1988). An investigation of differential prediction with the WISC-R. *Journal of School Psychology, 26*, 59-68.
- REYNOLDS, C. R. (1979). The invariance of the factorial validity of the Metropolitan Readiness Tests for blacks, whites, males, and females. *Educational and Psychological Measurement, 39*, 1047-1052.
- REYNOLDS, C. R. (1980). An examination for bias in a pre-school test battery across race and sex. *Journal of Educational Measurement, 17*, 137-146.
- ROSS-REYNOLDS, J., & RESCHLY, D. J. (1983). An investigation of item bias on the WISC-R with four sociocultural groups. *Journal of Consulting and Clinical Psychology, 51*, 144-146.
- SANDOVAL, J. (1979). The WISC-R and internal evidence of test bias with minority groups. *Journal of Consulting and Clinical Psychology, 47*, 919-927.
- SANDOVAL, J., & MILLE, M. W. (1980). Accuracy of judgments of WISC-R item difficulty for minority groups. *Journal of Consulting and Clinical Psychology, 48*, 249-253.
- SANDOVAL, J., ZIMMERMAN, I. L., & WOO-SAM, J. M. (1983). Cultural differences on WISC-R verbal items. *Journal of School Psychology, 21*, 49-55.
- SCARR, S. (1981). *Race, social class, and individual differences in IQ*. Hillsdale, NJ: Erlbaum.
- SCARR, S., & WEINBERG, R. A. (1983). The Minnesota adoption studies: Genetic differences and malleability. *Child Development, 54*, 260-267.
- SCHIELE, J. H. (1991). An epistemological perspective on intelligence assessment among African American children. *Journal of Black Psychology, 17*(2), 23-36.
- SHEPARD, L., CAMILLI, G., & AVERILL, K. M. (1981). Comparison of procedures for detecting test-item bias with both internal and external ability criteria. *Journal of Educational Statistics, 6*, 317-375.
- STIGLER, J., SCHWEDER, R., & HURT, R. (1990). *Cultural psychology*. New York: Cambridge University Press.
- VERNON, P. E. (1979). *Intelligence: Heredity and environment*. San Francisco: W. H. Freeman.
- WEINBERG, R. A., SCARR, S., & WALDMAN, I. D. (1992). The Minnesota transracial adoption study: A follow-up of IQ test performance at adolescence. *Intelligence, 16*, 117-135.
- WHITWORTH, R. H., & GIBBONS, R. T. (1986). Cross-racial comparison of the WAIS and WAIS-R. *Educational and Psychological Measurement, 46*, 1041-1049.
- WILLIAMS, R. L., & MITCHELL, H. (1991). The testing game. In R. L. Jones (Ed.), *Black psychology* (pp. 193-205). Berkeley, CA: Cobb & Henry.
- WURTZ, R. G., SEWELL, T., & MANNI, J. L. (1985). The relationship of Estimated Learning Potential to performance on a learning task and achievement. *Psychology in the Schools, 22*, 293-302.

SANDRA SCARR

CULTURE AND INTELLIGENCE The study of differences in intelligence among ethnically different populations, especially differences between Western and non-Western peoples, has a long history. The study of the relationship between culture and differences in intelligence is, however, a recent development. M. H. Segall and associates (1990) note that

Europeans or Westerners have long “felt a need to discover how non-Europeans think.” They believe that “primitive peoples” think differently and are intellectually inferior. These assumptions have influenced the study of intelligence among non-Western peoples and hindered investigators from examining the relationship between culture and intelligence. Nonetheless, the situation is changing with the emergence of cross-cultural psychology.

The traditional view of intelligence, which ignores culture, is a product of two factors. The first is Levy-Bruhl’s claim in 1910 that the thinking of non-Western peoples was “pre-logical.” By this he meant that non-Western peoples do not permit contradictions in their thinking and that they tend to equate themselves or human beings with animals and inanimate objects where the latter serve as totems. Levy-Bruhl based his conclusions on analysis of the religious beliefs of non-Western peoples, not on study of individuals. Under criticisms from anthropologists (e.g., Boas, 1911), he subsequently modified his theory (1948) but his speculations continue to influence students of intelligence to believe that the intelligence of non-Western peoples is qualitatively different and inferior.

The second factor is the theory of cultural evolution (Tylor, 1865). This theory asserted that human beings had progressed from savagery through barbarism to civilization. Different populations passed through the same stages of evolution but at different rates and some populations have not completed the evolutionary process because they have been “arrested” at some stages along the way (Segall et al., 1990). Psychologists applied this societal theory to individuals, proposing that individual development is a kind of linear unfolding process. They claimed that development from childhood to adulthood paralleled development from a primitive society stage to a civilized stage, that “ontogeny” recapitulated “phylogeny” (Hall, 1904). Some even suggested that the intelligence of some non-Western peoples was at the stage of children’s mental development (Segall et al., 1990).

Researchers no longer designate non-Western societies as “primitive”; instead, they distinguish between Western and non-Western, or “literate” and nonliterate societies. Nonetheless, some researchers still assume that the intelligence of non-Western or

nonliterate societies is qualitatively different from and probably inferior to that of Western people. Thus, some designate the intelligence of Western people as “open” and that of non-Western people as “closed” (Horton, 1967a, 1967b) or consider Western intelligence as “operational” and the intelligence of non-Western peoples as nonoperational (Hallpike, 1979). Similarly, the poor and minorities within Western societies are said to think at a “concrete” level, in contrast with the middle class, whose thinking is at an “abstract,” “conceptual,” and “logical” level (Jensen, 1969).

Contemporary study of intelligence among non-Western people shows a dramatic shift toward taking culture into account. The newer approach owes its heritage to the notion of the “psychic unity of mankind” proposed by Boas (1911; Wallace, 1961). The immediate impetus is to discover whether the stages and processes of cognitive development postulated by Jean Piaget are universal. Piaget’s “clinical method” is also appealing to cross-cultural researchers. This method involves “naturalistic” observation of younger children and “clinical” interviews with older ones. The interview technique is prevalent. Simply put, the researcher presents the child with a problem-solving situation, varying the same situation to probe the limit of the child’s ability. Both the observational and interview techniques require an extensive knowledge of the child’s cultural milieu and language, as well as a good rapport with the child (Dasen, 1977).

Cross-cultural psychologists have enhanced our understanding of cross-cultural differences in intelligence. Their findings have cast doubt on the idea of qualitative differences in intelligence among members of different cultures. They suggest that such differences as may exist derive from the ways cultures utilize the intelligence of their members. Thus, after reviewing available evidence, S. Scribner and M. Cole (1973, p. 553) conclude that “all cultural groups thus far studied have demonstrated the capacity to remember, generalize, form concepts, operate with abstractions, and reason logically.” However, cross-cultural studies continue to report that non-Western children lag two to seven years behind Western children in cognitive development if they have not been influenced by Western cultures (e.g., schooling). Some of the rea-

sons for cross-cultural differences in intelligence and for the difficulty in explaining the differences are conceptual, some are methodological. We deal with the conceptual issues first.

WHAT IS CULTURE?

Culture Defined. From the perspective of cultural anthropology, *culture* is a people's way of life and cannot be fully comprehended with the concept of "environment" or "ecology" (Ogbu, 1981). The culture of any population has four components:

1. Customary ways of behaving—activities such as making a living, expressing affection, raising children, responding to illness and to death, getting ahead in society, dealing with the supernatural
2. Codes or assumptions, expectations and emotions underlying those customary behaviors
3. Artifacts—things made by members of the population that have meanings for them
4. Institutions—economic, political, religious, and social—the imperatives of culture.

They form a recognizable pattern requiring competencies (including cognitive competencies) and customary behaviors (including "intelligent" behaviors as defined by the population) in a fairly predictable manner. Culture shapes the attributes of its members. People create, change, and pass on their culture to their children who in turn may change it. Culture also changes through external circumstances (Cohen, 1971; LeVine, 1967).

People feel, think, and behave in "cultural worlds," and each human population lives in a somewhat different cultural world. Culture is the framework within which its members see the world around them, interpret events in that world, behave according to acceptable standards, and react to their perceived reality. To understand the thoughts and behaviors of members of different populations it is necessary to understand their culture.

An example of a customary behavior in the United States is the ritual of caring for the mouth. Americans usually have shrines in their homes for daily mouth rituals (brushing their teeth) and occasionally they consult a "holy-mouth-man" (dentist), who specializes in the magical care of their mouths. The assumption

underlying this cultural behavior is the belief that the body houses two dangerous elements, debility and disease, which must be prevented from breaking out. Another example of cultural or customary behavior in the United States is the "stylin' out" of the black preacher through a special "code talk." The preacher's code talk is specialized to facilitate in-group feelings and to conceal black aspirations and feelings from the dominant white society. Although whites and blacks both "speak English," this behavior makes it difficult for whites to understand the language and style of blacks. Americans do not, of course, consciously analyze their cultural behaviors. For non-Americans to understand how Americans behave and why they behave as they do, however, requires more than a casual or short-term visit to the United States.

Cultural Differences. Cultural differences exist in the context of behaviors as well as in the underlying assumptions, rules, and meanings of these behaviors. For example, to most people in the United States, raising eyebrows means a surprise, but for people of the Marshall Islands in the Pacific Ocean, it signals an affirmative answer, and for the Greeks, it is a sign of disagreement. Cultures also differ in customary behaviors for social mobility. In Western societies, the middle-class strategy emphasizes individual competition based on ability or fate. Among the Lowland Christian Filipinos in the Philippines, an individual achieves and is expected to achieve social mobility through group cooperation. The Kanuri of Northern Nigeria seek social mobility through a patron-client relationship (Cohen, 1965).

Differences also exist in languages members of different cultures speak and in the way they use language to codify their environments and experiences. Therefore, speakers from two language groups may differ in the labeling and meaning of the same environment or experience. Some concepts that appear "natural" to members of one culture are not necessarily universal. Some concepts may be absent in a language because they are not used to code the group's environment, experience, or activities, that is, they are not a part of the group's functional adaptation. They are not absent because the people lack the requisite biological structures or genes, fail to teach them to their children, or because individuals or the group lag in "development."

For example, English speakers have several terms for ideas, objects, and behaviors associated with flying, such as fly (noun), fly (verb), pilot, airplane. Hopi speakers, on the other hand, have only one term for the idea and object, *fly*. English speakers have two terms for snow; Eskimos, for whom snow plays an enormous part in their cultural adaptation, have several terms. The Ibos of tropical Nigeria, for whom snow plays no part in their adaptation, have no word for it. English speakers distinguish between “blue” and “green,” and Navajo speakers lump “blue” and “green” together (Fishman, 1960).

Cultures differ in mathematical systems, concepts, and behaviors. The Western decimal or 10-system of numbers, for instance, is but one of several number systems in the world. Some Native American tribes have similar decimal or 10-systems. That is, they form their number words based on groupings of 10. Other Native American tribes have a 20-system, as do the Celts of Northwestern Europe, and Ainu of Northeastern Asia, the Yorubas and Ibos of Nigeria, and the Baganda of Uganda. Some Australian tribes and the Bushmen of Southwest Africa have a 2-system (Closs, 1986).

Differences in mathematical concepts and behaviors are illustrated when the Kpelle of Liberia are compared with Americans (Gay & Cole, 1967). The two groups are similar in arithmetic concepts because both classify things. But they differ because the Kpelle do not carry out such an activity explicitly. Neither do the Kpelle have concepts like “zero” or “number” in their counting system; and they do not have concepts for abstract operations like addition, subtraction, multiplication, and division, even though in their cultural mathematical behaviors, they add, subtract, multiply, and divide things. The Kpelle measure length, time, volume, and money like Americans but do not measure weight, area, speed, or temperature.

Why are there cultural differences? Cultural differences exist because populations differ in their cultural adaptations or their solutions to common human problems. Cultures differ in the way they solve common human problems, that is, in the means by which they make a living (economy and technology), govern themselves (politically), organize their domestic life for reproduction (family and childrearing/education),

manage their relationship with the supernatural (religion), and exchange ideas with one another (language and communication). Each domain both requires and promotes its own repertoire of competence, knowledge or cognition, assumptions, and emotions that support appropriate or “intelligent” behaviors. Populations differ in these solutions or cultural adaptations because they live in different physical and/or social environments and because of their different histories. Generally, members of a culture label these domains as well as label the functional competencies and their behaviors.

All cultural differences are not, however, alike. John Ogbu (1992) has distinguished two types of cultural differences in performance on intelligence studies. One is primary cultural differences, the differences that existed before two cultures came into contact. This type is found in the study of non-Western peoples and immigrants. The other, secondary cultural differences, is found among nonimmigrant minorities in urban industrial societies. Secondary cultural differences develop as responses of minorities to “oppression” by the dominant group and are usually oppositional. The two types of cultural differences will affect responses to psychological tasks in cognitive studies differently (Ogbu, 1988).

The broader view of culture is that it helps an individual become a contributing member of his or her social group. As children mature, they acquire their group’s customary behaviors, the thoughts and emotions that accompany and support such behaviors, the knowledge and meanings of cultural artifacts or symbols, the knowledge of societal institutions, and the practical and cognitive skills that make the institutions work. They acquire their group’s language and the means to communicate appropriately or intelligently; they learn to think, categorize, form concepts, generalize, label, remember, and reason logically in the manner sanctioned by their culture. These adaptations are attributes and behaviors valued in the group and fostered consciously and unconsciously by childrearing agents. They are not “out there” for students of intelligence or “short-term visiting researchers” to pick up and use or to discover them in test scores. The researcher should discover them before conducting a study.

WHAT IS INTELLIGENCE?

Conventional Definition. Neither conventional definition nor theory of intelligence sheds light on cultural differences in intelligence. Psychologists do not, of course, agree as to what “intelligence” is but many of them agree that it can be measured. Usually when they refer to “intelligence,” they mean the intelligence quotient (IQ) which is what they often measure (Jensen, 1969). Psychologists also do not agree as to what causes intelligence. They have debated for generations about what is more important in causing people to be intelligent—heredity (nature) or environment (nurture). Many hereditarians and environmentalists subscribe to “the ability theory” of intelligence. This theory states that intelligence is like a genealogical tree, with the generalized intelligence (the *g* factor) at the base; above it are specialized types of abilities (such as verbal, numerical, spatial-perceptual, memory, reasoning, and mechanical intelligence). Proponents of the ability theory believe that IQ tests tap some universal mental capacities. Therefore, the test scores of individuals and groups indicate their levels and types of mental abilities. The theory makes no allowance for the fact that test takers from different cultures may perceive the test items and testing situation differently or use different strategies from those intended by the testers or that verbal, numerical, and spatial-perceptual skills are valued differentially by various cultures.

Alternative Definitions. Two alternative conceptions of intelligence help to explain cross-cultural differences in intelligence. One view proposed by Vernon (1969) distinguishes among Intelligences A, B, and C. Intelligences A and B correspond to the geneticist’s distinction between the genotype and the phenotype. Intelligence A, the genotype, is the innate capacity children inherit from their parents. Intelligence A determines the limits of individuals’ intellectual development. For members of a population, Intelligence A represents their genetic potential for intellectual development. But a psychologist has no way to observe or measure Intelligence A directly (Vernon, 1969).

Intelligence B, the phenotype, is a product of both nature (genetic material) and nurture (environmental pressures and forces, or culture). It refers to everyday observed behavior of individuals considered intelligent or nonintelligent by members of the culture. Thus, In-

telligence B is culturally defined; hence, it varies from culture to culture even though the underlying capacities (i.e., Intelligence A) are apparently the same. For example, the behaviors which the middle class of Western urban industrial culture consider as intelligent are probably different from the behaviors which the Ibo subsistence farmers of Nigeria or subsistence Eskimo hunter-gatherers in Alaska value as intelligent behaviors. Western intelligence is a product of a historical adaptation to Western culture, whose technoeconomic and bureaucratic activities require and promote cognitive skills and strategies or intelligence characterized by grasping relations and symbolic thinking. These attributes have permeated to some extent school learning, activities at work and in daily life (Vernon, 1969). Note, however, that Western Intelligence B is not simply a matter of cultural definition; it also denotes the cognitive skills and strategies selected from a common species’ pool for adaptation to its specific environment and historical circumstances (Ogbu, 1978).

Intelligence B is not fixed. It can change when cultural change occurs (e.g., a move from rural to urban community, school attendance, or participation in advanced technology or economy). Such changes may make a person appear more intelligent or less intelligent relative to his or her peers. Similarly, the Intelligence B of a population may change because of significant changes in cultural activities. A good example is the change that occurs in the cognitive skills or intelligence of non-Western people when they participate in Western-type schooling (Cole & Scribner, 1974; Greenfield, 1965; Wagner & Stevenson, 1982).

The Intelligence B of the contemporary Western middle class is changing because of cultural change, namely, the emergence of computer technology and “high-tech” jobs with their emphasis on “cognitivism.” The cognitive skills required by the new computer technology include “precise definitions, linear thinking, precise rules and algorithms for thinking and acting” (Committee of Correspondence on the Future of Public Education, 1984). Middle-class people are acquiring these new cognitive skills by learning to use computers at home, at school, and at the workplace. Their schools are rewriting their curricula and modifying their instructional techniques to emphasize cog-

native skills compatible with computer thinking and high-tech know-how (Ogbu, 1988).

Intelligence C, measured intelligence or IQ, refers to those middle-class cognitive skills, or Intelligence B, usually sampled by IQ tests. Intelligence C is more limited than and differs from Intelligence B because the sampled cognitive skills for IQ tests are selected to predict scholastic performance or ability to perform other specific technological, economic, or societal tasks. That is, IQ represents part, but not all, of middle-class intelligence, from which it has been selected for specific purposes. Because IQ tests are made up of skills drawn from Intelligence B that are vital for solving specific problems associated with their culture (industrialization, bureaucracy, and urbanism), the skills are valued and emphasized in middle-class education and childrearing. Scores on IQ tests merely reveal how successful children are in acquiring the cognitive skills they will need to participate successfully as adults in jobs and other positions in their indigenous cultures.

When IQ tests based on this Western middle-class culture or Intelligence B are given to children of other cultures, their lower performance is probably predictable. The reason for the superior performance of non-Western children who have attended Western-type schools in comparison with the nonschooled peers, is that the schooled groups have acquired attributes of Western middle-class Intelligence B. The IQ test scores do not reveal the cognitive skills or Intelligence B that both groups use to solve the cognitive problems indigenous to their cultures (LeVine, 1970).

The distinction among Intelligences A, B, and C also allows us to speculate about another reason for cross-cultural differences in measured intelligence or IQ. Group differences in Intelligence B appear to derive primarily from cultural differences because populations do not necessarily differ in biological potential for intelligence. We begin to understand this by distinguishing maturational intellectual outcomes from cultural intellectual outcomes suggested by Ginsburg and Opper (1981). Jean PIAGET's idea of stages of cognitive development illustrate maturational outcomes. One maturational outcome is that of an individual reaching the stage of formal operational thinking. To reach this stage, a person's physical systems—the brain and central nervous system—must be developed fully or matured for the thought and language char-

acteristic of the formal-operational stage. No human population fails to reach this physical system maturity for operational thinking. Cross-cultural studies show, however, that populations vary in formal operational thinking (Dasen, 1977). The probable reason for the variation is that in some populations, such as Western societies, cultural tasks (e.g., bureaucratic and technological roles and schooling) require and promote a relatively high degree of formal operational thinking. As a result, formal-operational thinking is functional, valued and inculcated. By contrast, in some non-Western societies, cultural tasks do not call for much formal operational thinking, which apparently is not valued highly or implanted. Administering IQ tests loaded with formal operational tasks of Western Intelligence B to such populations predictably will yield lower scores.

Intelligence as a Functional System. S. Scribner and M. Cole (1973) proposed another formulation that may explain cross-cultural differences in intelligence. Following A. R. LURIA and Lev VYGOTSKY, they distinguish between (1) lower psychological processes, such as sensation and movement, and (2) higher psychological processes, such as voluntary memory, active attention, and abstract thought. They say that the latter are culturally organized into functional systems by members of culture on the basis of their historical, practical, and theoretical activities. Functional systems change when members of a culture undergo social or culture change so that the things they used to do and the way they used to do them change. Such changes in turn result in changes in the way they think or in their functional systems.

The introduction of Western-type formal schooling produces such changes. When schooling is introduced to a non-Western people, the latter begin to acquire new ways of using language (e.g., as an explicit tool for information exchange), "a scientific" approach to learning, a tendency to use generalized rules and verbal definitions, competence in use of symbols, such as numbers in math. They learn to learn out of context. All these changes result in the group's acquisition of a new *functional system*, or new intelligence. It is no wonder that in non-Western societies, children with some schooling perform better than their nonschooled peers on intelligence tests and on Western functional systems.

CROSS-CULTURAL DIFFERENCES IN INTELLIGENCE: THE ROLE OF CULTURE IN NON-WESTERN SOCIETIES

The idea that cross-cultural differences in intelligence test scores may arise because psychological tasks making up the tests are based on "alien cultures" receives some support from cross-cultural studies. These studies show that when tests are based on things people do or things familiar in their daily life, they tend to do well. Thus, after reviewing various studies of conservation tasks Cole and Scribner (1974) conclude that

there are indications that a number of specific experiential factors play an important role in performance on classification, psychological problem-solving tasks: familiarity with materials, opportunities presented by the environment for exploring spatial relationships, social contact with urbanized people, attendance at Western-type schools—all have been implicated as factors in the performance of one or more tasks.

Some studies show that performance on cognitive tasks is influenced positively by cultural experience. The kinds of activities in which members of a culture engage tend to stimulate the development and expression of those cognitive and perceptual skills necessary for competence in the culturally valued activities. One example is the study in Mexico by D. R. Price-Williams and colleagues (1969) in which experience with pottery making (and therefore manipulation of clay materials) resulted in superior performance in conservation tasks. W. Dennis's summary (1970) of scores on the Goodenough Draw-A-Man test from over forty different cultures is another example. His comparative analysis shows a strong relationship between a group's involvement with representational art and children's test scores. Dennis found almost no sex differences, except for two high-scoring groups, namely, Native Americans in the Southwest and villagers in rural art centers in Japan. In these two groups the sex differences were in reverse order. Among the Japanese villagers, where women rather than men were highly involved with art work, girls scored higher than boys. In the American Southwest, Native American groups, where men rather than women were involved with art

work, boys scored higher than girls. P. J. Zho (cited in Serpell, 1982) has also reported a similar influence of sex-typed art activities on test scores for rural Ngoni children in central Africa.

Still another example comes from P. R. Dasen's study (1974) of cognitive development among Australian aborigines and white Australian children. Dasen found that white Australian children, like their European and American peers, develop logical mathematical concepts *before* they develop spatial concepts; but among the aborigines, the order of development is reversed. Dasen's reasonable interpretation is that the aborigines develop spatial concepts earlier than logical-mathematical concepts because the former are more important for their nomadic hunting and gathering economy or adaptation.

Finally, cultural influences on mathematical skills and behaviors show up in tests. G. B. Saxe and J. Posner have reported that two West African cultures whose economic activities differ, also develop different numerical skills (Posner, 1982; Saxe and Posner, 1983). One group, the Dioula, consists of merchants whose principal economic activities require extensive use of numerals; their culture not only values mathematical skills but also provides children with many opportunities to practice and develop these skills. The second group, the Baoule, consists of subsistence farmers whose major economic activities do not depend on numerical skills and who, therefore, do not value these skills strongly enough to stress them for their children. Children from the two groups who had not been to Western-type schools performed differently when tested for mathematical skills. Dioula children, the merchant group, were superior. Saxe and Posner (1983, p. 303) summarize the difference between the two groups by saying that the unschooled Dioula children "adopt more economic strategies than the unschooled Baoule, the agricultural group, counterpart. In particular, they use a greater number of memorized addition facts and regrouping by tens ($7 + 5 = 10 + 2$), compared to the Baoule."

METHODOLOGICAL ISSUES

Some cross-cultural differences in intelligence arise from methodological difficulties. These differences are a particularly important consideration when interpret-

ing the reports that non-Western children tend to lag behind their Western peers by two to seven years in intellectual development. Cross-cultural psychologists are well aware of some of their methodological difficulties (Cole et al., 1971; Cole & Scribner, 1974; Segall et al., 1990). Cole and his associates remark that the experimental approach of the psychologist is not adequate and suggest that it should be combined with the ethnographic approach of the anthropologist into a new methodology—ethnographic psychology.

Segall and colleagues (1990) discuss methodological problems with respect to what to measure, sampling, and test administration. Among the measurement problems is the fact that the behavior measured by the psychologist may have different meaning in the culture. Consider, for example, the different meanings of raising eyebrows in three cultures discussed earlier. Sampling problems include representativeness and the tendency to use easily accessible subjects. Test-administration problems also arise from a lack of effective communication with subjects, because the researcher may not be fluent in the local language.

There are other methodological issues. Most cultural studies focus on the concrete operational stage and occasionally on the preoperational stage. One reason for focusing on the concrete operational stage is that its cognitive contents, such as conservation and classification, are measurable through standardized tasks. Researchers resort to measurement and quantification, which they criticize and want to avoid.

Three additional problems arise. One is that the actual technique employed in field studies is not “clinical” in the way it was practiced by Piaget. Reports from cross-cultural studies do not indicate that researchers know enough about the children in their samples to help them make reasonably good interpretations of the children’s responses to the problem-solving situations (Kamara & Easley, 1977).

Except in few instances where cross-cultural researchers are natives of the culture of their subjects or are anthropologists, researchers are not usually fluent in the native language. Nonetheless, in order to conduct a proper “clinical” interview, the researcher and the subject must be fluent in the same language. Fluency permits the researcher to be more aware of the subtle distinctions between the reality perceptions of the subjects and their reliance on social cues or

simply guessing the correct answer. This problem is not solved adequately by using translators or interviewing the subjects in the language of the researcher. To solve this problem researchers must be willing and able to invest sufficient time and effort to learn the language of their subjects, as a good ethnographer does.

Lack of familiarity with the local culture makes truly clinical interviews difficult in three ways. First, researchers do not know enough about local cultures to recognize those real-life situations in the local cultures that require and promote the particular cognitive skills being studied. Some activities embodying the intellectual skills investigated by researchers, such as classification, are usually present in some areas of local cultures. A researcher can use the knowledge of such activities and situations to design the study and to distinguish the problem of performance from that of competence in interpreting subjects’ responses.

Second, subjects are given unfamiliar materials in the psychological tasks and this adversely affects their responses. The literature indicates that subjects do relatively well in the few instances in which they are familiar with the “test” materials. For example, local water vessels in some cultures can be used instead of Western beakers in the study of conservation of liquid. Elaborate conservation may be found in many areas of local cultures, such as in subsistence and physical environment (Ogbu, 1982).

Third, cultural barriers from the researchers’ lack of knowledge of the subtleties of interpersonal relationships within the culture interfere with clinical interviews. Researchers often do not understand the cultural norms regarding adult–child relationships and behaviors. This affects their ability to interpret accurately the subjects’ behaviors in test situations.

CONCLUSION

Do different members of different populations think differently? Is there a “primitive mentality”? Do group differences in IQ test scores derive from genetic differences, environmental differences, or cultural differences? These are old and unresolved questions about group differences in intelligence.

Cross-cultural studies suggest that all investigated populations have the capacities for the various cogni-

tive processes that constitute intelligence. They can categorize, form concepts, generalize, operate with abstraction, remember, and reason logically. How they do these things appears to be determined by culture. Therefore, actual cognitive behaviors vary from culture to culture. As a result, any test of cognitive behavior, such as an IQ test, designed to suit members of one population or culture will discriminate almost inevitably against members of other cultures. There is no "culture-free" intelligence test.

Conventional IQ tests measure culturally valued cognitive behaviors of the Western middle class. They discriminate inevitably against tribal and peasant peoples and against minorities within Western societies. But tribal and peasant peoples as well as immigrant minorities change their cognitive behaviors when they begin to participate in Western-type education and technoeconomic systems. This shift tends to eliminate or reduce cross-cultural differences in intelligence. The same is true of immigrant minorities, some of whom surpass their Western peers.

The people who have most difficulty with IQ tests and other forms of cognitive tasks are involuntary or nonimmigrant minorities. This difficulty arises because their cultures are not merely different from that of the dominant group but may be in opposition to the latter. Therefore, the tests acquire symbolic meanings for these minorities which cause additional but as yet unrecognized problems. It is more difficult for them to cross cognitive boundaries.

(See also: CROSS-CULTURAL VARIATIONS IN INTELLIGENCE; EASTERN VIEWS OF INTELLIGENCE.)

BIBLIOGRAPHY

- ADJEI, K. (1977). Influence of specific maternal occupation and behavior on Piagetian cognitive development. In P. Dasen, (Ed.), *Piagetian psychology: Cross-cultural contributions* (pp. 227–256). New York: Gardner Press.
- AL-FAKHRI, S. (1977). The development of the concept of speed among Iraqi children. In P. Dasen, (Ed.), *Piagetian psychology: Cross-cultural contributions* (pp. 203–226). New York: Gardner Press.
- BOAS, D. (1911). *The mind of the primitive man*. New York: Macmillan.
- CLOSS, M. P. (ED.). (1986). *Native American mathematics*. Austin: University of Texas Press.
- COHEN, R. (1965). Some aspects of institutionalized exchange: A Kanuri example. *Cahiers D'Études Africaines*, 5, 353–369.
- COHEN, Y. A. (1971). The shaping of men's minds: Adaptations to the imperatives of culture. In M. L. Wax, S. Diamond, & F. O. Gearing (Eds.), *Anthropological perspectives on education* (pp. 19–50). New York: Basic Books.
- COLE, M. (1975). An ethnographic psychology of cognition. In R. W. Brislin, S. Bochner, & W. J. Lonner (Eds.), *Cross-cultural perspectives on learning* (pp. 157–175). New York: Halsted Press.
- COLE, M., GAY, J., GLICK, J., & SHARP, D. W. (1971). *The cultural context of learning and thinking*. New York: Basic Books.
- COLE, M., & SCRIBNER, S. (1974). *Culture and thought: A psychological introduction*. New York: Wiley.
- COMMITTEE OF CORRESPONDENCE ON THE FUTURE OF PUBLIC EDUCATION. (1984). *Education for democratic future: A manifesto*. New York: Author.
- DASEN, P. R. (ED.). (1974). The influence of ecology, culture and European contact on cognitive development in Australian aborigines. In R. P. Dasen (Ed.), *Culture and cognition*. London: Methuen.
- DASEN, P. R. (ED.). (1977). Introduction. In P. R. Dasen (Ed.), *Piagetian psychology: Cross-cultural contributions* (pp. 1–25). New York: Gardner Press.
- DENNIS, W. (1970). Goodenough scores, art experience, and modernization. In I. Al-Issa & W. Denis (Eds.), *Cross-cultural studies of behavior* (pp. 134–152). New York: Henry Holt.
- EDGERTON, R. (1982). Another look at culture and mental retardation. In M. J. Begab, H. C. Haywood, & H. L. Garber (Eds.), *Psychosocial influences in retarded performance* (Vol. 1, pp. 309–324). Baltimore, MD: University Park Press.
- FISHMAN, J. A. (1960). Systematization of the Whorfian hypothesis. *Behavioral Science*, 5, 323–339.
- GAY, J., & COLE, M. (1967). *The new mathematics and an old culture: A study of learning among the Kpelle of Liberia*. New York: Henry Holt.
- GINSBURG, H. P., & OPPER, S. (1981). *Piaget's theory of intellectual development: An introduction*. Englewood Cliffs, NJ: Prentice-Hall.
- GREENFIELD, P. (1965). On culture and conservation. In J.

- S. Bruner, R. R. Oliver, & P. M. Greenfield (Eds.), *Studies in cognitive growth* (pp. 225–256). New York: Wiley.
- HALL, G. S. (1904). *Adolescence: Its psychology, and its relation to physiology, anthropology, sociology, sex, crime, religion, and education* (2 vols.). New York: Appleton.
- HALLPIKE, C. R. (1979). *The foundations of primitive thought*. Oxford: Clarendon Press.
- HEBB, D. O. (1949). *The organization of behavior*. New York: Wiley.
- HOLLOS, M., & RICHARD, F. (1993). Gender-associated development of formal operations in Nigerian adolescents. *Ethos*, 21(1), 24–52.
- HORTON, R. (1967a). African traditional thought and Western science, Part 1: From tradition to science. *Africa* 37(1), 50–57.
- HORTON, R. (1967b). African traditional thought and Western science, Part 2: The “closed” and “open” predicament. *Africa*, 37(2), 155–187.
- JENSEN, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1–127.
- KAMARA, A. I. (1971). *Cognitive development among school-age Themne children of Sierra Leone*. Unpublished doctoral dissertation, University of Illinois, Urbana.
- KAMARA, A. I., & EASLEY, J. A., JR. (1977). Is the rate of cognitive development uniform across cultures?—A methodological critique with new evidence from Themne children. In P. R. Dasen (Ed.), *Piagetian psychology: Cross-cultural contributions* (pp. 26–63). New York: Gardner Press.
- KIMINYO, D. M. (1977). A cross-cultural study of the development of conservation of mass, weight, and volume among Kamba children. In P. R. Dasen (Ed.), *Piagetian psychology: Cross-cultural contributions* (pp. 64–88). New York: Gardner Press.
- KIRK-FITZGERALD, L. (1970). *Cognitive development among Ga children: Environmental correlates of cognitive growth rate within the Ga tribe*. Unpublished doctoral dissertation, University of California, Berkeley.
- LAVE, J. (1977). Tailor-made experiments and evaluating the intellectual consequences of apprenticeship training. In *The Quarterly Newsletter of the Institute for Comparative Human Development*, 1(2), 1–3.
- LEVINE, R. A. (1967). *Dreams and deeds: Achievement motivation in Nigeria*. Chicago: University of Chicago Press.
- LEVINE, R. A. (1970). Cross-cultural study in child psychology. In P. Mussen, (Ed.), *Carmichael's manual of child psychology*. New York: Wiley.
- LEVY-BRUHL, L. (1910). *Les fonctions mentales dans les sociétés inférieures*. Paris: Alcan.
- LEVY-BRUHL, L. (1949). *Les carnets de Lucien Levy-Bruhl*. Paris: Presses Universitaires de France.
- OGBU, J. U. (1978). *Minority education and caste: The American system in Cross-cultural perspective*. New York: Academic Press.
- OGBU, J. U. (1981). Origins of human competence: A cultural-ecological perspective. *Child Development*, 52, 413–429.
- OGBU, J. U. (1982). *Do the Poka “see hunger all the time”? The role of wild plants in Poka diet*. Paper presented at the 81st annual meeting of the American Anthropological Association, Washington, DC, December, 1982.
- OGBU, J. U. (1987). Cultural influences on plasticity in human development. In J. J. Gallagher & C. T. Ramey (Eds.), *The malleability of children* (pp. 155–169). Baltimore: Paul H. Brookes.
- OGBU, J. U. (1988). Human intelligence testing: A cultural-ecological perspective. *National Forum: A Journal of the Phi Kappa Phi*, 66(1), 23–29.
- OGBU, J. U. (1992). Understanding Cultural Diversity and Learning. *Educational Researcher*, 8(21), 5–14.
- OKONJI, O. M. (1971). A cross-cultural study of the effects of familiarity on classificatory behavior. *Journal of Cross-Cultural Psychology*, 2, 39–49.
- POSNER, J. K. (1982). The development of mathematical knowledge in two West African societies. *Child Development*, 53, 200–208.
- PRICE-WILLIAMS, D. R. (1961). A study concerning concepts of conservation of quantities among primitive children. *Acta Psychologica*, 18, 293–305.
- PRICE-WILLIAMS, D. R., GORDON, W., & RAMIREZ, M. (1969). Skill and conservation: A study of pottery-making children. *Developmental Psychology*, 1, 769.
- SAXE, G. B., & POSNER, J. (1983). The development of numerical cognition: Cross-cultural perspectives. In H. P. Ginsburg (Ed.), *The development of mathematical thinking* (pp. 291–317). New York: Academic Press.
- SCRIBNER, S., & COLE, M. (1973). Cognitive consequences of formal and informal education. *Science*, 182, 553–559.
- SEGALL, M. H., DASEN, P. R., BERRY, J. W., & POORTINGA, Y. H. (1990). *Human behavior in global perspective: An introduction to cross-cultural psychology*. Boston: Allyn & Bacon.
- SERPELL, R. (1982). *Intellectual handicap in a cross-cultural perspective*. Paper presented at the Sixth International Congress of the International Association for Cross-Cultural

- Psychology, University of Aberdeen, Scotland, July 20, 1982.
- TAYLOR, R. B. (1988). *Cultural ways: A concise introduction to cultural anthropology*. Prospect Heights, IL: Waveland Press.
- TYLOR, E. B. (1865). *Primitive culture*. London: John Murray.
- VERNON, P. E. (1969). *Intelligence and cultural environment*. London: Methuen.
- WAGNER, D. A., & STEVENSON, H. W. (EDS.). (1982). *Cultural perspectives on child development*. San Francisco: W. H. Freeman.
- WALLACE, A. F. C. (1961). *Culture and personality*. New York: Random House.
- WERNER, H. (1948). *Comparative psychology of mental development*. New York: Science Editions.

JOHN U. OGBU

D

DECISION MAKING AND JUDGMENT

Decision making is the process by which people make choices among alternatives. For example, a person who moves to a new city may need to decide how to get to work each day. This process begins by identifying the relevant alternatives from among the expansive set of potential modes of transport. In the case of commuting to work, the person might consider driving alone, car pooling, taking the bus, or walking as the most relevant among the set of potential transportation alternatives available (rejected from consideration might be boating, roller-skating, etc.). In some cases, this may involve generating new options not currently being used; biking to work may appeal to the decision maker, though no one else uses this option.

The decision maker then needs to evaluate each of the considered alternatives. How does each of them serve the several goals important to the person? Should the selected choice be ecologically sound? Then driving alone is a nonpreferred alternative. However, are time and convenience high priorities? Then driving alone becomes the most attractive option. Is exercise considered to be desirable? Then walking and biking take precedence. Is flexibility an important consideration? Then car pooling may be a poor alternative.

As the example illustrates, several dimensions are likely to be important in the decision, so that a transportation option that is ecologically sound, quick, convenient, flexible, and enhances physical fitness would

be adopted with confidence by the decision maker. Because few decisions have one alternative that meets all needs, the decision maker may need to set priorities among several important goals, with the expectation of having to make some compromises.

Thus, the process of decision making requires identifying and comparing alternatives, each of which varies on a number of relevant dimensions. For any given decision, some optimal process might be specified that fully incorporates all relevant information and would result in a judgment most likely to meet the decision maker's goals. Following this best-judgment rule is called "optimizing," with the resultant judgment pattern being fully rational.

Decades of research in human information processing show that humans are limited in how much information they can manage at one time. How do people make such complex decisions with an information-processing capacity that is so much more limited than the demands of the situation? Studies show that they systematically simplify the decision.

Overwhelmed by decision complexity, the judge may use a simplified decision strategy, one that gets by in most cases but may have liabilities that result in occasional erroneous conclusions. These mental shortcuts are called "heuristics," and the pattern of decisions resulting from such strategies is termed "bounded rationality." Psychological studies of human judgment and decision making explore the strategies

employed to simplify the decision-making environment and the extent to which those strategies may compromise optimality. Several aspects of the decision-making process that have proven central are the strategies involved in comparing choice alternatives, ways people reason about statistical relationships, and their confidence in their own judgments.

COMPARING CHOICE ALTERNATIVES

One aspect of the complexity of decision making is managing the informational and computational aspects of comparing the alternatives being considered. How will the alternatives be evaluated on the relevant dimensions, and how will that information be combined to reach a final conclusion? Studies show several common shortcuts.

One common way in which decisions are simplified is by limiting the information considered. Studies show that decision makers base their choices on only a fraction of the potentially relevant information, with that proportion decreasing as the total amount of available information increases (e.g., Payne, 1976). People are especially insensitive to missing information and are unlikely to detect categories of information that should be supplied before reaching a conclusion. Decision makers have difficulty even imagining the full set of possible outcomes that should be considered, further shortcutting decision complexity.

Once the information has been gathered, a complete comparison of alternatives would assess each alternative (e.g., biking, car pooling, driving alone) on each of the relevant dimensions (e.g., cost, flexibility, safety) and combine the set of assessments to compare the transportation alternatives fully. Yet, studies show that people use a variety of rules to combine this information into some summary judgment, many of which considerably simplify the process.

One simple approach would be elimination by aspects (Tversky, 1972). Using this strategy, the decision maker selects a criterion value on one dimension and rejects any alternatives that do not meet this condition. In the transportation decision, the commuter may deem any option that takes longer than thirty minutes as unacceptable, resulting in rejection of walking, taking the bus, and biking to work. Flexibility may

be the next aspect to be considered, at which point car pooling would be rejected, leaving driving alone as the one surviving alternative and hence, the transportation mode of choice. Note that while the decision has been considerably streamlined in this way, issues of ecological impact, exercise, and convenience, although important to the decision maker, were never considered. The same judgment strategy that began by evaluating the alternatives on those dimensions might result in an entirely different conclusion.

Evidence identifies a variety of rules by which humans compare choice alternatives, with no single rule holding across all domains. Rather, decision makers seem to select among strategies according to the decision conditions (e.g., time pressure) and the particular choice being made (e.g., how important is the outcome?). Decision makers may even shift strategies in the course of a single decision, perhaps using an elimination-by-aspects approach to reject clearly inappropriate alternatives and then making a more detailed comparison of a few attractive possibilities.

STATISTICAL REASONING

Decision making is further complicated by uncertainty about how the alternatives will meet one's goals. For example, riding the bus to work may be inexpensive, but how likely is it that the bus will be on time? What is the chance that the fares will go up? A car pool may be an attractive alternative, but how likely is it that the other riders will do their share of the driving? Will they be prompt? Uncertainty and risks like these enhance the complexity of the decision maker's task in finding the alternative that best serves his or her goals.

Uncertainty in a decision environment requires an understanding of probability, likelihood, and randomness, all aspects of statistical reasoning. Unfortunately, human intuitions about statistical relationships are often faulty, in some cases with serious implications for decision adequacy. Studies show several judgment heuristics that lead to problems in probabilistic reasoning.

Availability. Using the availability heuristic, an individual estimates event likelihoods by the ease with which examples come to mind (Tversky & Kahneman,

1974). For example, are words in the English language more likely to begin with a *k* or have a *k* in the third position? There are actually three times as many words with *k* in the third position, but most people conclude that *k* is more common in the first position, since such words are easier to recall (e.g., kangaroo, kick).

This misleading strategy may have serious consequences. For example, differential availability of examples may underlie people's tendencies to fear unlikely but vivid risks (e.g., assault, death by fire) even as they underestimate risks of higher likelihood that actually take more lives (e.g., traffic accidents). If protective measures are based on this reasoning, people may leave themselves most exposed to the greatest threats to health and safety (e.g., failure to use seat belts) and devote most of their efforts to avoiding risks they are unlikely to encounter.

The availability heuristic may also underlie belief perseverance; that is, examples that support one's beliefs may be easier to recall than contrary cases, leading the individual to conclude that the belief is well supported by the evidence. For example, individuals biased against car pools may find cases of car pools that did not work out (coworkers unfriendly, unreliable, etc.) easy to recall, while those which worked effectively may be buried in memory. Thus, a real-world experience base full of variety can be differentially available, supporting false beliefs.

Representativeness. Using the representativeness heuristic, a person estimates an event's likelihood by the extent to which it represents the typical features of its category. For example, consider the relative likelihood of two coin-toss sequences in which heads (H) and tails (T) come up H-T-H-T-T-H and H-H-H-T-T-T. Most people will conclude that the former is more likely than the latter sequence, since it has the haphazard mixture of heads and tails they expect of randomness. In fact, the two sequences are actually equally likely.

The representativeness heuristic may underlie people's inclination to be excessively suspicious of "runs" in random series, which are then attributed to non-random processes. For example, basketball fans commonly believe in "hot" streaks in shooting, times when an athlete consistently makes baskets. This is seen as a temporary characteristic of the athlete, one that may

pass with no warning, but should be taken advantage of while it lasts. Investigation of actual shooting percentages show that the "hot hand" is an illusion (Gilovich, 1991). A player is no more likely to make a basket after making the previous basket than after missing the previous one. The fact that baskets are sometimes made in streaks seems aberrant from people's expectations about random variation; hence, they attribute it to the player.

Cumulative Risk. In many cases, probabilistic information is combined in the decision-making process. Consider a risk people take regularly, such as that of being injured in a car accident. The risk of injury associated with each car trip is quite low, but since most people take several car rides per day for 365 days per year over the many years of their lives, the cumulative risk of being in at least one car accident over a whole lifetime is quite high. Yet, research shows that people underestimate the long-term implications of chronic exposure to low-level daily risk (Shaklee & Fischhoff, 1990). Such reasoning may underlie people's reluctance to take such precautions against a variety of daily risks as using seat belts to minimize driving risk, condoms to prevent exposure to AIDS, life vests while boating, and safety glasses for machine work.

Problems in understanding cumulative risk may be a special case of a more general heuristic, anchoring and adjustment. People often make a cumulative estimate by taking a starting value and then making sequential adjustments in that value to reach a final answer. Since the size of those adjustments are typically insufficient, the final estimate is too close to the starting value. For example, people were asked to estimate the product of either $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$ or $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$. Although the two problems have the same product, the group estimating the solution to the former problem gave estimates of more than four times the size of the estimates of latter group of people (median estimates of 2,250 versus 512), and each group greatly underestimated the true answer (40, 320). Having started with a higher value, the people in the first group more closely approximated the true answer, yet their series of insufficient adjustments still left them far off the mark in final estimate. Similarly, in the case of risk, people may be aware that cumulative

risk increases with length of exposure, but their minor increments from the starting short-term risk value result in an underestimate of the long-term risk.

Framing. One property of statistical relationships is that alternative expressions of the same relationship should be treated equivalently, but human judgment is reliably affected by form of statistical expression. For example, people show different preferences for disaster-relief programs when they are described in terms of lives lost rather than in terms of lives saved, or for meat products when described in terms of percent fat or percent lean.

Related evidence shows that preference for gambles depends on how people think of the gamble. For example, questions about the value of a gamble (e.g., how much you would be willing to pay for a lottery ticket) lead people to focus on the money aspects of the gamble (i.e., how much the bet pays). With attention on the payoff, those individuals neglect the probability of winning, showing excessive interest in a low-likelihood gamble. The same people switched their preference to a high-likelihood payoff when they rated gambles for attractiveness, a judgment that increased attention to the probability of winning. Such preference reversals were even shown by gamblers in a Las Vegas casino (Lichtenstein & Slovic, 1973). While the actual expected value was constant, alternative ways of thinking about the gamble resulted in nonoptimal betting behavior even when the stakes were real.

CONFIDENCE

One way in which people recognize uncertainty in the decision environment is by expressing less than perfect confidence in their conclusion. Is decision-maker confidence a reliable indicator of likely judgment accuracy?

Calibration. The extent to which judgment accuracy matches judgment confidence is called "calibration." For those cases in which the decision maker is 90 percent confident, 90 percent of the conclusions should be accurate; a set of judgments at the 50 percent confidence level should be correct half of the time. Studies of calibration show that confidence and accuracy are correlated but overconfidence is the

norm (Lichtenstein, Fischhoff, & Phillips, 1977). Confidence was especially unwarranted for high-confidence values. For example, answers with confidence judgments at the 90 percent level proved to be accurate only 75 percent of the time. Further studies suggest that people's confidence is bolstered by considering the evidence in favor of their chosen alternative. Calibration is improved when people are also asked to consider evidence in favor of the rejected alternative (Koriat, Lichtenstein, & Fischhoff, 1980).

Hindsight. Predicting events is a common activity. How likely is it that the Democratic candidate will win the presidential election? What is the probability the home team will win the baseball game if the relief pitcher comes in? What is the likelihood that today's romance will turn into a long-term relationship? When the events have yet to take place, uncertainty seems greatest.

But once the outcome is determined, what happens to that uncertainty? Research suggests that people tend to exaggerate the extent to which they could have predicted the outcome all along, a pattern termed the hindsight bias (Fischhoff, 1975). Before the event, people may be quite uncertain about which of several alternative outcomes would prevail, but once the results are in, they exaggerate the predictability of that outcome. Thus, the Democrat's victory seems like it was assured from the first, the manager should have known better than to switch the pitcher so late in the game, and once the wedding date is set, the romance seems as if it had been predetermined from the first date.

Illusion of Control. Research suggests that people show inappropriate confidence even when events are known to be random (Langer, 1975). In one case, people who made a random draw of a card in competition against a nervous opponent were more confident of a win than were those who faced a confident opponent. In another study, people who selected their own lottery tickets were more confident their number would be drawn than were those who had no choice about which ticket to take. In yet another demonstration, those who could practice a response with a randomly determined outcome were more confident they would have a winning outcome than were those with no practice time. The study sug-

gests that all of these effects are the product of the illusion of control people experience when the attributes of skill tasks (e.g., practice, decision making) are brought into settings in which the outcome is due to pure luck.

DECISION COMPETENCE AND JUDGMENT DEMANDS

The research literature in the psychology of judgment identifies a number of ways in which human capacity or strategy is ill suited to the demands of the decision-making environment. Information-processing demands may exceed mental capacity, necessitating mental shortcuts. Or people may approach the problems with intuitively based rules that violate formal statistical properties. Finally, whatever the conclusion, people are likely to have inflated confidence in its accuracy.

These strategies and heuristics may lead to a conclusion that will be adequate in many cases, but in others, the judgment is likely to be wrong, sometimes with serious consequences. Research on human judgment and decision making has helped clarify the liabilities of those judgment strategies.

This research offers a contrast to much of the research on human intelligence, where individual differences have been the central topic of concern. That is, some individuals bring better reasoning capacities than others to the problem-solving situation. In contrast, judgment research has emphasized the demands of the decision-making environment, demands even the best of decision makers may have problems meeting.

Yet, people need to make decisions on a regular basis. How can they improve their performance? Research shows that many of these heuristics and shortcuts are resistant to training. More successful have been attempts to develop aids to allow decision makers to specify which information is relevant and what their priorities are, and then the judgment aid manages the information load and computational aspects, which prove so challenging to human limitations.

BIBLIOGRAPHY

FISCHHOFF, B. (1975). Hindsight \neq foresight: The effect of outcome knowledge on judgment under uncertainty.

Journal of Experimental Psychology: Human Perception and Performance, 1, 288–299.

GILOVICH, T. (1991). *How we know what isn't so*. New York: Free Press.

KAHNEMAN, D., & TVERSKY, A. (1982). The psychology of preference. *Scientific American*, 246, 160–173.

KORIAT, A.; LICHTENSTEIN, S.; & FISCHHOFF, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory*, 6, 107–118.

LANGER, E. (1975). The illusion of control. *Journal of Personality and Social Psychology*, 32, 311–328.

LICHTENSTEIN, S., FISCHHOFF, B., & PHILLIPS, L. (1977). Calibration of probabilities; The state of the art. In H. Jungerman & G. deZeew (Eds.), *Decision making and change in human affairs*. Dordrecht, Netherlands: D. Reidel.

LICHTENSTEIN, S., & SLOVIC, P. (1973). Response-induced reversals of preference in gambling: An extended replication in Las Vegas. *Journal of Experimental Psychology*, 101, 16–20.

NISBETT, R., & ROSS, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.

PAYNE, J. (1976). Task complexity and contingent processing in decision making: An information search and protocol analysis. *Organizational Psychology and Human Performance*, 16, 366–387.

SHAKLEE, H., & FISCHHOFF, B. (1990). The psychology of contraceptive surprises: Cumulative risk and contraceptive effectiveness. *Journal of Applied Social Psychology*, 20, 385–403.

SLOVIC, P. (1987). Perception of risk. *Science*, April, 280–285.

SLOVIC, P. (1990). Choice. In D. Osherson and E. Smith (Eds.), *Thinking*. Cambridge, MA: MIT Press.

TVERSKY, A. (1972). Elimination by aspects: A theory of choice. *Psychological Review*, 79, 281–299.

TVERSKY, A., & KAHNEMAN, D. (1974). Judgment under uncertainty; Heuristics and biases. *Science*, September, 1124–1131.

TVERSKY, A., & KAHNEMAN, D. (1981). The framing of decisions and the psychology of choice. *Science*, January, 453–458.

HARRIET SHAKLEE

DEMENTIAS See BRAIN, PATHOLOGIES OF THE

DEVELOPMENT, COGNITIVE Cognitive development involves the growth of thought processes. Although important changes in thinking occur over the human life span, the rate of change seems especially dramatic during infancy, childhood, and adolescence. For this reason, this entry focuses on the changes in thinking that occur during this early period.

Although we all know what thinking is, defining it cleanly is quite difficult. No sharp boundary separates activities that involve thinking from activities that do not. Listing examples of thinking turns out to be easier and more fruitful than formally defining the term. The activities that come to mind most quickly when we think about thinking refer to higher mental processes: solving problems, reasoning, conceptualizing, planning, reorganizing, classifying, symbolizing, reading, writing, calculating, and so on. Thinking also involves more basic processes, processes that even young children perform skillfully: using language, perceiving objects and events, and remembering, to cite three. Still other activities are gray areas that might or might not be viewed as thinking: being socially skillful, having a keen moral sense, feeling appropriate emotions, and so on. The qualities in this last group clearly involve thought processes but also involve many other, non-intellectual processes.

Cognitive development raises questions of theoretical interest to philosophers and psychologists and questions of practical interest to parents and educators. How do infants see the world—in basically the same way as older children and adults, or in fundamentally different ways? What are the roles of maturation and experience? Is it harmful to try to teach children to think in advanced ways at early ages? Where do children's ideas come from? Is children's thinking basically the same as that of ignorant adults, or does their thinking differ above and beyond their lack of knowledge? Why do some children learn more readily than others? Does thinking progress through qualitatively distinct stages or is cognitive growth a gradual and continuous process? Perhaps most fundamental, how does cognitive change occur?

Some of the most interesting aspects of cognitive development involve the dramatic differences between the reasoning of young children and that of adults. R. DeVries (1969) described one of the most remarkable of these differences. She examined 3- to 6-

year-olds' understanding of the difference between appearance and reality. The children were presented an unusually good-natured cat named Maynard and were allowed to pet him. When the experimenter asked what Maynard was, all of the children knew he was a cat. Then the experimenter put a mask with the face of a fierce dog on Maynard's head and said, "Look, it has the face of a dog. What is this animal now?"

Many of the 3-year-olds thought that putting the mask on Maynard had transformed him from cat to dog. They refused to pet him and said that under his skin he had a dog's bones and a dog's stomach. In contrast, most 6-year-olds knew that a cat could not turn into a dog and that the mask did not change the animal's nature.

Such dramatic differences in reasoning by children of different ages have led some to propose that cognitive development involves progress through a series of qualitatively distinct stages. The most prominent such stage theory is that of the Swiss psychologist Jean PIAGET. Piaget proposed that the first two years after birth are spent in the sensorimotor stage; the years from 2 to 7, in the preoperational stage; the years from 7 to 12, in the concrete operations stage; and the rest of life, in the formal operations stage. Within Piaget's theory, cognitive development in the sensorimotor stage primarily involves development of skill in interacting motorically with the world; development in the preoperational period involves acquisition of language and mental imagery; development in the concrete operations period involves acquisition of ability to represent transformations; and development of formal operations involves acquisition of ability to reason in a purely abstract fashion.

Piaget's theory remains the single most prominent approach to cognitive development. The reasons are that it covers the full age range from birth to adolescence, includes a vary broad range of types of thinking, and includes numerous surprising and interesting findings about children's thinking (for example, his classic conservation studies, in which he found that most 5-year-olds believe that pouring water from one glass to a differently shaped glass changes the amount of water). In general, it conveys a good feel for the types of changes that occur in the course of cognitive development. Nevertheless, the theory also has been shown to be limited in a number of ways. It depicts

between-stage changes as being discontinuous, where most evidence available in the early 1990s indicates cognitive change to be a basically continuous process. It also fails to recognize that a number of early-developing competencies that children acquire before their stage classification would suggest that such acquisitions are possible, and it underestimates the contribution of experiential factors to development. Still, it remains the single most encompassing theory of cognitive development and provides a good overview of major trends that occur from birth through adolescence.

Probably the largest difference between Piaget's depiction of cognitive development and what has emerged from research in the early 1990s is that infants are far more cognitively competent than Piaget (or anyone else) believed until quite recently. The impressive capabilities uncovered by recent investigations can be illustrated in the context of distance perception. Philosophers have long speculated about how people are able to perceive an object's distance from themselves. Some, such as George Berkeley, an associationist philosopher of the eighteenth century, concluded that the only way in which infants could accurately perceive distance was by crawling around the environment and associating how objects looked with how much movement was required to reach them. Yet Carl Granrud (1992) demonstrated that on infant's first full *day* out of the womb, they already perceive which objects are closer and which are farther away. Clearly, infants perceive depth before they can move around the environment.

Infants also possess more impressive conceptual skills than previously realized. For example, one of Piaget's best-known claims was that until 8 or 9 months of age, infants lack object permanence, the knowledge that objects continue to exist even when they cannot be seen. Piaget inferred that infants lack such knowledge from observations that 6- and 7-month-olds who were playing with a ball would make no effort to retrieve it if the ball was placed under an opaque cup. This was not due to a lack of interest or motoric incapacity; the same infants would retrieve the ball if it was placed under an otherwise identical transparent cup (Bower & Wishart, 1972). The observation suggested that, for infants, out of sight was quite literally out of mind.

The research of Renée Baillargeon (1992), however, has shown that much younger infants actually represent objects not visible to them. Infants of 3 and 4 months of age show surprise when a moving object appears to travel through a location that had been occupied by another stationary object, even when the infants can no longer see the original object (the physically impossible event is an illusion accomplished with mirrors). The infants' surprise indicates that they expect the original object to continue existing at its original location and to preclude the moving object from traveling through that location.

These are just two of the many late-twentieth-century discoveries of heretofore unsuspected competence in infants and young children. Long before they go to school, they have surprising knowledge of such fundamental concepts as time, space, number, and causality; they even possess simple theories of mind (Wellman, 1990).

The deficiencies of Piagetian theory have stimulated the development of several alternative theories. The most prominent of these are information-processing theories of development. These theories vary in their particulars, but all share the view that cognitive development is due in large part to some combination of four types of development: development of basic processes, strategies, content knowledge, and metacognitive understanding. Some information-processing theories, such as that of Case (1985), are quite close to Piagetian theory in postulating stages of development, but place greater emphasis on the development of basic processes such as working memory and central conceptual structures. Other information-processing approaches, such as that of Keil (1989), view development as the formation of increasingly sophisticated theories that combine associative and causal knowledge. Yet others envision cognitive development as a self-modifying computer program, in which an innate kernel of learning capacities operates on experience to produce development.

A key unifying factor among information-processing theories is a focus on the mechanisms that produce development. Rapid progress is being made on this issue at both the physiological and the cognitive levels (for a review, see Siegler, 1989).

One type of progress involves the identification of changes in the brain during infancy. Synaptogenesis,

the formation of connections among neurons within the brain, appears to be a large contributor to development. The number of synapses within numerous parts of the brain follows a distinctive developmental course, in which there is an initial overproduction and later pruning of synaptic connections. Synaptic connections are most numerous during the early postnatal period. For example, the average number of synaptic connections in one part of the brain (the third layer of the middle frontal gyrus) grows from 10,000 to 100,000 between birth and 12 months and continues to increase until age 2. In the next 5 years, more synapses are pruned than formed, leading the density of synapses to decrease to adult levels by age 7. In general, from 6 months to 7 years, the synaptic density in children's brains appears to exceed that in the brains of adults (Huttenlocher, 1979). A number of investigators have linked synaptic density to ability to learn and form representations and have argued that the early high density may play a critical role in early cognitive attainments (e.g., Goldman-Rakic, 1987).

The subsequent pruning of unneeded connections appears to improve the efficiency of performance without interfering with the by-now-acquired cognitive competencies. Which particular synapses are pruned depends on experience. Normal experience at the normal time results in neural activity that maintains typical connections; abnormal experience or lack of experience at the usual time results in connections being preserved that usually would be pruned. Such a process allows both efficient acquisition in normal environments and reasonable adaptation to abnormal circumstances. In particular, the genes provide a rough outline of the eventual form of the process, thus allowing quite rapid acquisition under typical circumstances. Unusual environments or physical deficiencies, however, lead to unusual neural activity, which creates alternative neural organizations that are adaptive, given the circumstances.

A large number of cognitive mechanisms contribute to development. One of the most widely influential is strategy choices. People can approach almost any task in multiple ways. In order to reach optimal performance, they must choose among these strategies adaptively. For example, even young children tend to know a variety of strategies for solving simple addition problems: counting on their fingers, using better-known

problems as reference points, retrieving answers from memory, guessing, and so on. To proceed as accurately and efficiently as possible, they must choose from among these strategies which to use on each problem and how to adapt to varying situational demands for speed and accuracy.

Even 5-year-olds make such choices in surprisingly adaptive ways. Again, this can be illustrated in the context of simple addition (Siegler, 1986). Given a choice of retrieving an answer from memory or counting on their fingers, 5-year-olds generally retrieve answers to the easier problems from memory and solve the harder ones by counting. This allows them to proceed quickly and accurately on the easy problems and accurately, albeit less quickly, on the harder ones. They also adapt their counting to the type of problem. For example, on problems such as $2 + 9$, with a large difference between the addends, they generally count on from 9 and simply say "9, 10, 11." They also shift strategies primarily when the strategies they are using yield incorrect answers and when they know strategies that are likely to be more accurate.

The strategy-choice process contributes to development in several ways. One is to produce knowledge leading to increasingly adaptive choices, as children learn which types of strategies work best on which types of problems. Another contribution comes in building up knowledge of problems and thus helping them in the future to retrieve answers to the problems. The strategy choices also help children discover new strategies by demonstrating the properties that useful strategies in the domain must have. Through studying synaptogenesis, strategy choice, and a variety of other mechanisms, researchers are making rapid progress on the single most fundamental issue regarding cognitive development: How does change occur?

(See also: INFANCY; INFANT TESTS AS MEASURES OF EARLY COMPETENCE; INTERVENTIONS, INFANT AND PRESCHOOL; IQ GAINS OVER TIME; PIAGETIAN THEORY OF DEVELOPMENT.)

BIBLIOGRAPHY

- BAILLARGEON, R. (1992). The object concept revisited: New directions in the investigation of infants' physical knowl-

- edge. In C. Granrud (Ed.), *Visual perception and cognition in infancy*. Hillsdale, NJ: Erlbaum.
- BOWER, T. G. R., & WISHART, J. R. (1972). The effects of motor skill on object permanence. *Cognition*, 1, 165–172.
- CASE, R. (1985). *Intellectual development: Birth to adulthood*. New York: Academic Press.
- DEVRIES, R. (1969). Constancy of generic identity in the years three to six. *Society for Research in Child Development Monographs*, 34 (Whole No. 127).
- GOLDMAN-RAKIC, P. S. (1987). Development of cortical circuitry and cognitive function. *Child Development*, 58, 601–622.
- GRANRUD, C. (1992). *Visual perception and cognition in infancy*. Hillsdale, NJ: Erlbaum.
- HUTTENLOCHER, P. R. (1979). Synaptic density in human frontal cortex: Developmental changes and effects of aging. *Brain Research*, 163, 195–205.
- KEIL, F. C. (1989). *Concepts, kinds, and cognitive development*. Cambridge, MA: MIT Press.
- SIEGLER, R. S. (1986). Unities in strategy choices across domains. In M. Perlmutter (Ed.), *Minnesota symposium on child development* (Vol. 19). Hillsdale, NJ: Erlbaum.
- SIEGLER, R. S. (1989). How domain-general and domain-specific knowledge interact to product strategy choices. *Merrill-Palmer Quarterly*, 35, 1–26.
- WELLMAN, H. M. (1990). *Children's theories of mind*. Cambridge, MA: MIT Press.

ROBERT S. SIEGLER

DEVIATION QUOTIENT (DQ) The earliest method of computing a person's measurable level of intelligence was a test that yielded a score called the mental age (MA), introduced by Alfred BINET and Théophile Simon in their 1905 Binet-Simon Scale. In 1912, the mental age score was converted into an intelligence quotient (IQ) score by William Stern, who divided an individual's MA by that person's actual chronological age (CA) and multiplied the result by 100. LEWIS TERMAN of Stanford University retained this method of computing IQ scores in his expanded and refined 1916 and 1937 Stanford-Binet adaptations of the Binet-Simon Scale.

David WECHSLER of New York City's Bellevue Hospital introduced his Wechsler-Bellevue Scale for assessing the intelligence of adults in 1939; the earlier

scores involving an MA and CA in computing an IQ had been based on testing children. In place of that older IQ concept, Wechsler substituted the deviation quotient (DQ), an index obtained from converting an individual's raw test score into an IQ number that expresses the deviation from the mean, which he set arbitrarily at 100. Wechsler wanted to assess each examinee's measurable ability directly against that of others within the same age group, using as the reference point only the scores of those tested in that age group. This provided an IQ score derived from the distribution of only those tested in each age group, such that 50 percent in each age group earned a score of 100 or above and 50 percent earned a score of 100 or below. The usable range of such IQ scores is from 50 to 150. In current usage, the terms *DQ* and *IQ* are used interchangeably, although most professionals prefer the older, generic term, *IQ*.

BIBLIOGRAPHY

- KAUFMAN, ALAN S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn & Bacon.
- MATARAZZO, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence: Fifth and Enlarged edition*. New York: Oxford University Press.

JOSEPH D. MATARAZZO

DIALECTICAL THINKING The term *dialectical thinking* has become popular since the 1970s. It refers to thinking that acknowledges change, diversity, and contradiction. This form of thinking is contrasted to thinking that attempts to arrive at and assumes the existence of answers to questions and resolutions of problems that are definitely true or false or definitely desirable or undesirable and that do not vary as a result of individuals' backgrounds or other factors (Riegel, 1973). Dialectical thought is not based on such right/wrong logic. Instead, in dialectical thinking we usually feel that whether or not something is true is a matter of time and context. We also feel that what is true or right cannot be resolved in an either/or fashion. Rather, we realize that both sides of a duality need to be considered and that, paradoxically, both sides taken together and in interaction are more true than either side alone.

Interest in dialectical thinking has been revived because there is a sense that our culture for a long time was under the hold of a form of thinking in which we feel that questions have one single correct answer and problems have one single correct solution that holds for everybody, despite variations in circumstances (see Labouvie-Vief, 1994). Such thinking is based on an either/or logic in which only one solution can be right. This belief is the cause of many dualisms, such as the opposition of mind versus body, reason versus emotion, and so forth. Dialectical thinking rejects such either-or, right-wrong logic. Instead if we think dialectically we feel that whether something is true is a matter of time, context, or perspective, and that the truth lies in diversity and tension rather than the removal of contradiction. Thus, for example, if we ask whether the cause of an illness is mental or physical, we reject an either/or answer. Instead, we may assert that mental and physical causes combine to produce illness (Reich, 1992).

Many individuals have suggested that dualistic, either/or logic is only one possible way of thinking. This realization has emerged in many branches of modern science and philosophy (Baynes, Bohman, & McCarthy, 1987). It has also been important in research on the thinking of adults, who are less likely to think in terms of either/or answers than children and adolescents. However, the ideal of dialectical thinking is actually quite ancient. The ancient Greek philosopher Heraclitus (who lived even earlier than Socrates), for example, taught that flux, by which he meant diversity, opposites, conflict, and change, was the basic principle of the universe, and good thinkers had to recognize these qualities. Thus, he believed that good thinking was based on paradoxical opposites. This paradoxical model of thought also was predominant in Chinese and Indian thinking. Its basic premise was expressed in the sixth century B.C. by the Chinese philosopher Lao-tse: "Words that are strictly true seem to be paradoxical. . . . Gravity is the root of lightness; stillness is the ruler of movement" (Fromm, 1956, p. 63). In more modern times, the notion of thinking in terms of paradoxical opposites was revived by the German philosopher Hegel (1770–1831), who first used the term *dialectical* for this kind of thinking. In Hegel's dialectic, a statement implies its opposite, and this

contradiction leads to a resolution in a new statement. This Hegelian notion has had a profound influence in modern thinking, and now we tend to believe that whether or not something is true cannot be discussed without considering time and context (Griffin, 1988).

RESEARCH ON DIALECTICAL THINKING

Because of the increased interest in dialectical thinking, it is not surprising that many psychologists have begun to study this form of logic and to determine how it changes as a function of age. The general belief is that dialectical thinking is difficult for children or adolescents, who think in terms of dualisms and right-wrong opposition. It may be easier, however, for adults, and perhaps it is most characteristic of the thinking of people we consider wise.

Several kinds of tasks have been used in research to study dialectical thinking. One of the earliest was by Perry (1968), who examined learning and decision-making processes in college students. Perry observed that students move through several levels. The first of those is dualistic: Here students assert that learning consists of acquiring facts that are either right or wrong, and that there are no gray areas. The next level is relativist: Students compare alternative viewpoints and understand that facts are not necessarily right or wrong but take on different meanings depending on one's viewpoint. Perry believed that this awareness of relativism can leave the individual hanging without any ability to judge. However, confident judgment returns as the individual realizes that we can commit ourselves even though there is no absolute way of knowing.

Perry observed that these levels also generalize to how individuals understand the role of authority in making decisions. Dualists believe that authorities have absolute knowledge about what facts are right and wrong; relativists believe that there are no facts that are right or wrong in an absolute way; those committed despite relativism understand that even though we may not know what is right or wrong in an absolute way, we nevertheless can make well-informed (even though potentially incorrect) decisions.

Kitchener and King (see Kitchener & Brenner, 1990) extended Perry's work by using a task in which

individuals are confronted with different ways of interpreting the same state of affairs. For example, in one task individuals were asked which account of creation was correct, that reported in the Bible or that taught by science. Dualists sided with one account; they thought this account was infallible and self-evident, while the other account was believed to be wrong. Some individuals, however, move on to a way of thinking in which these accounts are not necessarily contradictory; however, these individuals are comfortable with the idea that we cannot know for certain, but that often we are left with educated guesses.

Since Perry, a number of researchers have developed similar tasks and have shown that as individuals develop beyond youth and adolescence and into mature adulthood, they become less dualistic and better able to think in terms of change over time, comparisons of contrasting viewpoints, and the relationship of thinking to context. Such changes in thinking can have implications for many everyday kinds of tasks, for example, how individuals make sense of the opposing accounts of parties in a conflict (Kuhn, 1991), how they make sense of diversity in religion (Fowler, 1981), how they integrate the tension between the viewpoints on the self and others (Blanchard-Fields, 1986), or how they reason about the nature of their emotions (Labouvie-Vief, 1992).

One form of dialectical thinking that has been studied is the ability to understand that opposites can stand in a complementary rather than a contradictory relationship (see Reich, 1992). For example, in the nineteenth century, physicists thought that light had either the properties of a wave or of matter, and tried to find out which was the true description. Later, however, physicists proposed that such descriptions can form opposing though complementary ways of describing physical reality. Such thinking is of great importance in everyday life, as when we ask whether something is due to biology or culture, whether our behavior is determined by free will or by circumstances out of our control, or whether something is "in the mind" or "in the body." Broughton (1980) and Reich (1992) did research on such questions and found that older individuals are better able to think in terms of complementarity. For example, Reich asked children, adolescents, and adults whether a pianist had achieved

her level of skill as a result of her genetic endowment or of her diligent practice. A child aged 8 claimed that "inborn is important!" An adolescent aged 15 suggested that either practice or endowment could be the cause. But an adult aged 66 asserted that no either/or answer was possible.

Many studies have claimed that the ability to think dialectically, or in terms of complementarities, develops in adulthood. However, research suggests that this form of thinking may not be a separate form because it is highly related to measures that involve cultural knowledge (often called crystallized intelligence), such as measures of verbal ability. It is also not certain that this thinking is characteristic of mature adults, because it often emerges much earlier, sometimes in adolescence. In adulthood, research suggests that whether or not somebody thinks dialectically is actually not related to their chronological age. Instead, it seems to depend on the kind of teaching context individuals are exposed to. The degree to which an individual is emotionally open or closed also may be a factor. Finally, historical time also is an important factor, since dialectical thinking became more popular in the late twentieth century.

BIBLIOGRAPHY

- BASSECHES, M. A. (1984). Dialectical thinking as a metasystematic form of cognitive organization. In M. L. Commons, F. A. Richards, & C. Armon (Eds.), *Beyond formal operations* (pp. 216–238). New York: Praeger.
- BAYNES, K., BOHMAN, J., & MCCARTHY, T. (1987). *Philosophy: End or transformation?* Cambridge, MA: MIT Press.
- BLANCHARD-FIELDS, F. (1986). Reasoning on social dilemmas varying in emotional saliency: An adult developmental perspective. *Psychology and Aging, 1*, 325–333.
- BROUGHTON, J. M. (1980). Genetic metaphysics: The developmental psychology of mind-body concepts. In R. W. Rieber (Ed.), *Body and mind* (pp. 177–221). New York: Academic Press.
- COMMONS, M. L., RICHARDS, F. A., & ARMON, C. (1984). *Beyond formal operations: Late adolescent and adult cognitive development*. New York: Praeger.
- FOWLER, J. W. (1981). *Stages of faith: The psychology of human development and the quest for meaning*. San Francisco: Harper & Row.

- FROMM, E. (1951). *The forgotten language*. New York: Grove Press.
- GRIFFIN, D. R. (1988) (Ed.), *The re-enchantment of science: Post-modern proposals*. Albany, NY: State University of New York Press.
- KITCHENER, K. S., & BRENNER, H. G. (1990). Wisdom and reflective judgment: Knowing in the face of uncertainty. In R. Sternberg (Ed.), *Wisdom: Its nature, origin, and development* (pp. 212–229). Cambridge: Cambridge University Press.
- KUHN, D. (1991). *The skills of argument*. New York: Cambridge University Press.
- LABOUVIE-VIEF, G. (1992). A neo-Piagetian on adult cognitive development. In R. L. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 197–228). New York: Cambridge University Press.
- LABOUVIE-VIEF, G. (1984). *Psyche and Eros: Mind and gender in the life course*. New York: Cambridge University Press.
- PERRY, W. G. (1968). *Forms of intellectual and ethical development in the college years*. New York: Holt, Rinehart & Winston.
- REICH, K. H. (1992). Definitionen der Komplementarität. In E. P. Fischer, H. S. Herzka, & K. H. Reich (Eds.), *Widerspruechliche Wirklichkeit* (pp. 18–28). Germany: Piper.
- RIEGEL, K. F. (1973). Dialectical operations: The last period of cognitive development. *Human Development*, 16, 346–370.

GISELA LABOUVIE-VIEF

DIFFERENTIAL ABILITY SCALES The Differential Ability Scales (DAS) is a battery of individually administered subtests measuring mental abilities and school achievement among children aged 2 ½ through 17 years. It can be used by school and clinical psychologists and other trained professionals who need to assess a child's strengths and weaknesses as well as the child's overall level of intellectual ability. The DAS was designed especially to gather information that may be useful when working with children who have learning difficulties.

Several characteristics differentiate the DAS from other ability batteries for children. One is that the DAS distinguishes between *core* subtests that strongly measure an overall, general ability and *diagnostic* subtests that measure largely independent abilities such as

memory and speed of mental processing. Also, the DAS emphasizes the measurement of specific rather than global abilities. It measures an increasing number of distinct abilities as children get older. The DAS is administered in a way that tries to have each child take only those items and subtests that are at the right level of difficulty for that child.

THE COGNITIVE BATTERY

Seventeen of the twenty DAS subtests measure cognitive abilities (i.e., mental abilities involved in thinking, learning, and perceiving), and three measure school achievement. The cognitive subtests are divided into a preschool-age battery and a school-age battery. Eight of the cognitive subtests are used only at the preschool level, six are given only at the school-age level, and three are included at both levels. None of the cognitive subtests require the child to read or write, and only one of them (Speed of Information Processing) requires the child to work quickly.

Preschool Level. The colorful preschool-age subtests provide many materials such as toys and blocks for the child to manipulate. Some require verbal skills, while others require no language ability. Administration takes about twenty-five to sixty-five minutes, depending on the child's age and how many of the optional subtests are given.

The youngest preschoolers, those aged 2 years, 6 months ("2-6") through 3-5, take four *core* subtests whose scores are combined to give an overall General Conceptual Ability (GCA) score. They may also take two diagnostic subtests that are interpreted individually. Children aged 3-6 through 5-11 take six core subtests that provide both a GCA score and cluster scores in verbal ability and nonverbal ability. Five diagnostic subtests are available for these older preschool-age children.

The following are the subtests at the preschool level, and the ages at which each is most often given. Most of these subtests may also be given at older ages, typically up through age 7-11. The core subtests are

Naming Vocabulary (ages 2-6 through 5-11): naming an object shown in a full-color picture; part of the verbal score;

Verbal Comprehension (ages 2-6 through 5-11): manipulating toys or shapes or pointing to pictures in

response to spoken instructions from the examiner; part of the verbal score;

Picture Similarities (ages 2-6 through 5-11): selecting a picture that is similar to or related to a target picture; part of the nonverbal score;

Block Building (ages 2-6 through 3-5): copying a design or structure using wooden blocks;

Copying (ages 3-6 through 5-11): copying a design using pencil and paper; part of the nonverbal score;

Pattern Construction (ages 3-6 through 5-11): copying a pattern using multicolored squares or cubes; part of the nonverbal score; and

Early Number Concepts (ages 3-6 through 5-11): counting, recognizing numerals, understanding numerical concepts, and doing simple mental arithmetic using pictorial items.

The diagnostic subtests are

Recall of Digits (ages 3-0 through 5-11): repeating a string of digits spoken by the examiner;

Recognition of Pictures (ages 3-0 through 5-11): pointing to pictures that were seen before from among a larger set of pictures;

Recall of Objects (ages 4-0 through 5-11): naming from memory as many as possible of a set of twenty objects that were shown before; and

Matching Letter-Like Forms (ages 4-0 through 5-11): pointing to a letterlike shape that is exactly the same as a target shape.

School-Age Level. The school-age cognitive battery has six core subtests that give both a GCA score and cluster scores for verbal ability, nonverbal reasoning ability, and spatial ability (skill in visualizing shapes and patterns). There are also three diagnostic subtests. Administration takes about forty to sixty-five minutes. The subtests are all given at ages 6-0 through 17-11. The core subtests are

Word Definitions: giving an oral definition of a word spoken by the examiner; part of the verbal score;

Similarities: telling how three words spoken by the examiner are similar to one another; part of the verbal score;

Matrices: pointing to an abstract design that completes a logical pattern of designs; part of the nonverbal reasoning score;

Sequential and Quantitative Reasoning: drawing a design that fits into a logical sequence of designs, or saying a number that fits a pattern of numbers; part of the nonverbal reasoning score;

Pattern Construction (see preschool level): part of the spatial score; and

Recall of Designs: drawing an abstract design from memory; part of the spatial score.

The diagnostic subtests are

Recall of Digits (see preschool level);

Recall of Objects (see preschool level); and

Speed of Information Processing: quickly marking in each of many rows the circle containing the most small boxes, or the largest numeral.

THE ACHIEVEMENT BATTERY

The school achievement tests give a quick (fifteen to twenty-five minutes.) assessment of a school-age child's level of accomplishment in three areas. In the Word Reading test the child reads aloud words of increasing difficulty. In Spelling, the child writes words spoken by the examiner. The Basic Number Skills test consists primarily of an arithmetic computation worksheet, and also includes numeral-recognition items for the younger children and word problems for the older children. All three achievement tests were normed on the same sample as the cognitive battery so that meaningful comparisons between cognitive and achievement scores can be made.

DAS'S THEORETICAL BASIS

Although the DAS is consistent with a number of theories of human abilities, it is not based on any single theory. Its subtests were designed to give distinctive information about many ability domains, including some (such as speed of information processing) that reflect fairly recent research. Thus, the DAS is most compatible with theories, such as those of Thurstone (1938) and of Horn and Cattell (Horn, 1988), that emphasize multiple dimensions of cognitive ability.

The way in which a test battery is organized reflects its theoretical perspective on abilities. Two aspects of the DAS's organization are unusual and have theoretical implications. First, unlike most cognitive test batteries that combine the scores from all their subtests

into a single overall score, the DAS includes in its overall General Conceptual Ability (GCA) score only those subtests (the core subtests) that strongly reflect a general ability dimension. These core subtests measure reasoning with words, numbers, or figures; verbal knowledge; and the ability to solve visual-spatial problems. The other subtests, those that measure memory, perceptual matching, and information-processing speed, are kept separate from the overall composite. Although these latter abilities are important, excluding them from the GCA makes the GCA a focused, relatively pure measure of the general factor.

Figure 1 illustrates the organization of the school-age cognitive battery. The subtests (shown along the bottom row) measure abilities at their most specific level. The six core subtests form the base of a hierarchy that leads up to GCA. Pairs of core subtests define higher-level dimensions of ability (verbal, nonverbal reasoning, and spatial). These higher-level dimensions combine further to define the most general dimension, GCA. The remaining diagnostic subtests are not part of the hierarchy, but are nevertheless important because they give unique information about other strengths and weaknesses.

The second unusual feature of the DAS's organization is that it includes an increasing number of ability clusters (such as verbal ability or spatial ability) as children get older. Data collected during DAS development suggest that children's cognitive abilities become more differentiated (or, at least, more capable of being measured differentially) as their general intellectual

level develops. At the youngest ages all the subtests seem to reflect one broad ability, but at older ages children show more distinctive patterns of strengths and weaknesses.

What is the nature of GCA? The DAS does not use the term *intelligence* in describing it. The *DAS Handbook* (Elliott, 1990, p. 20) defines GCA as "the general ability of an individual to perform complex mental processing that involves conceptualization and the transformation of information." This ability is very close to the *g* factor that was proposed by Spearman (1927) and that is usually found to be the dimension that batteries of diverse cognitive tasks have in common. This dimension is defined more narrowly than is the concept of intelligence, a term that often is used to refer to a wide array of abilities and traits that contribute to the person's general functioning in everyday life. In other words, the more general concept of intelligence includes, but is not limited to, *g*.

For some users, the DAS's lack of a strong link to a particular theoretical model of human abilities may be a weakness. Tests that are not closely tied to a theory are limited in their ability to contribute to basic knowledge about the nature of cognitive abilities, and also do not have an existing theoretical basis for applying test results to educational or other problems. The DAS's main objectives are pragmatic rather than theoretical. In the *DAS Handbook*, Elliott expresses his view that the science of educational psychology has not developed to the point of being able to explain learning problems or offer theory-based recommen-

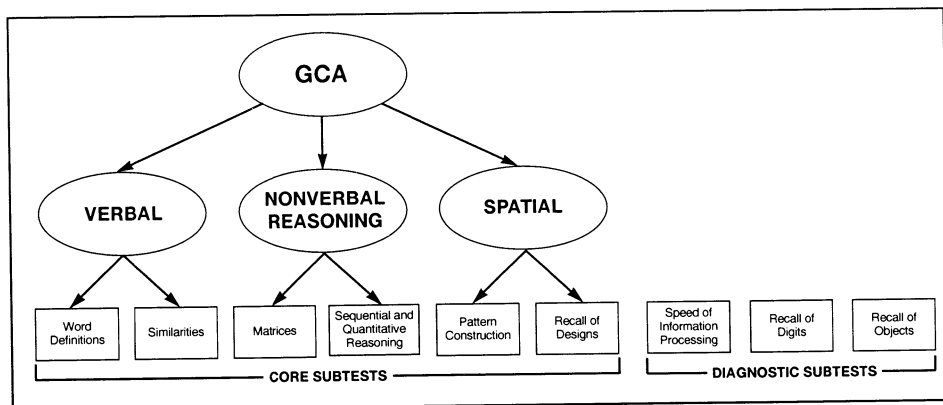


Figure 1

Structure of the DAS school-age cognitive battery

dations for correction, and that, therefore, tests should not tie themselves too closely to specific theories.

HISTORY AND DEVELOPMENT

The DAS is a major revision of the British Ability Scales (BAS), published in England in 1979 (Elliott, Murray, & Pearson, 1979). In 1965 the British Department of Education and Science had begun a project to create a British intelligence test for children. Colin Elliott of the University of Manchester's Department of Education became director of the project in 1973.

Elliott guided the development of the BAS along two basic principles. One was to measure a large number of distinct abilities to support differential diagnosis, with the assessment of *intelligence* (i.e., general functioning in everyday life) being less important. The second was to use a new method of test analysis, item response theory, to ensure that each subtest measured a single ability.

The development of the DAS, also under Elliott's authorship, followed these same principles. In addition to removing content that was specifically British, the DAS incorporated some major changes. Four subtests were added to expand the coverage of abilities at certain ages (particularly nonverbal measures for preschool-age children), and six of the BAS subtests were dropped. Administration procedures based on item-response theory were further developed to allow the examiner to select items that would be most appropriate to the individual child's ability level. The structure of core and diagnostic cognitive subtests was created.

Between 1987 and 1989 the DAS was normed in the United States on a sample of 3,475 children matched on appropriate demographic characteristics to 1988 U.S. Census data. The battery was published in 1990.

DISTINCTIVE FEATURES

A distinguishing feature of the DAS is the degree to which its subtests measure distinct abilities. The numerical measure of this characteristic, called *subtest specificity*, is high for the DAS in comparison with other similar instruments. High specificity provides a sound basis for identifying a child's strengths and weaknesses,

and that is why it was emphasized during subtest design. For example, the Speed of Information Processing (SIP) subtest is very accurate but is only weakly related to the GCA. Children at the same level on the GCA will have widely varying SIP scores. Therefore, the SIP subtest contributes unique information about the child.

A second unusual feature of the DAS is that it promotes *out-of-level* testing, that is, giving test material that matches the child's ability level rather than matching only his or her age. Items or entire subtests that are too easy or too difficult for most children of a particular age can be given if the child is low or high in ability. For example, almost all of the subtests in both the preschool and school-age batteries were normed at ages 5 through 7, so children in this transition age range can be given either level. Methods are provided whereby an older child whose development is seriously delayed can take the preschool-level battery and obtain a GCA score. Throughout the battery, the items in each subtest are organized so that each child takes only a group of items that are appropriate in difficulty, rather than having to take a large number of easier and more difficult items.

Because of its features for out-of-level testing, the DAS measures at relatively low ranges of ability. On most cognitive batteries the overall IQ (or its equivalent) can be measured in a range from about 40 to 160; the DAS's GCA score goes down to 25 for children aged 3-9 and older. In some cases, this extended range requires administering subtests normally used with younger children. The range also reflects the fact that the preschool subtests include very easy items suitable even for low-ability preschoolers.

TECHNICAL CHARACTERISTICS AND INTERPRETATION

The DAS norm sample, that is, the sample of cases used in test development that serves as a comparison group for interpreting each individual child's scores, contains 175 children for each 6 months of age from 2-6 through 4-11, and 200 per year of age from 5-0 through 17-11. The sample is evenly divided between boys and girls, and closely matches the 1988 U.S. population in race/ethnicity, geographical region, and parents' education. About 600 additional African-

American and Hispanic children were tested to provide large samples for detecting (and removing) possible bias in items and subtests.

Scores on the GCA and the ability clusters (such as verbal ability) are highly accurate, and statistical analyses show that many of the subtests also have relatively small amounts of measurement error. *Reliability* is an index of accuracy that ranges from 0.0 (for a completely inaccurate test) to 1.0 (perfect accuracy). The DAS cognitive subtests have an average reliability of .84, the achievement test reliabilities average .90, and the GCA reliability is .92 to .95, depending on age. *Test-retest* studies, in which children were tested again after about a month, also suggest that scores from a single administration are fairly accurate.

The GCA score has been found to agree (correlate) closely with the overall scores from the familiar Wechsler and Stanford-Binet intelligence tests. Like the scores from those instruments, the DAS's GCA and its verbal ability and nonverbal reasoning ability cluster scores are fairly good predictors of school achievement, as measured both by achievement tests and by teachers' grades.

Because the DAS was designed to measure profiles of strengths and weaknesses, the manual provides a great deal of guidance on how to identify important differences between scores. These differences can be the basis for hypotheses about strong or weak underlying processes, and those hypotheses may be helpful in designing different teaching strategies for individual children.

There is still much to be learned about the causes of strengths and weaknesses in cognitive processes, and about the roles played by these processes in children's learning problems. Therefore, interpreting a score profile is not a straightforward, simple process. Research evidence in general suggests that there are numerous causes of learning difficulties. Studies using the DAS and BAS, such as those by Kercher and Sandoval (1991) and by Tyler and Elliott (1988), have found learning-disabled children to cluster into subgroups with distinct profiles of high and low scores. Like other cognitive instruments, the DAS is a tool for research as well as a source of information for diagnosis.

(See also: INDIVIDUAL TESTS.)

BIBLIOGRAPHY

- ELLIOTT, C. D. (1990). *Differential Ability Scales: Introductory and technical handbook*. San Antonio, TX: The Psychological Corp.
- ELLIOTT, C. D., DANIEL, M. H., & GUITON, G. W. (1991). Preschool cognitive assessment with the Differential Ability Scales. In B. A. Bracken (Ed.), *The psychoeducational assessment of preschool children* (2nd ed.) (pp. 133-153). Boston: Allyn & Bacon.
- ELLIOTT, C. D., MURRAY, D. J., & PEARSON, L. S. (1979). *British Ability Scales*. Windsor, England: National Foundation for Educational Research.
- HORN, J. L. (1988). Thinking about human abilities. In J. R. Nesselrode & R. B. Cattell (Eds.), *Handbook of multivariate psychology* (2d ed.) (pp. 645-865). New York: Academic Press.
- KERCHER, A. C., & SANDOVAL, J. (1991). Reading disability and the Differential Ability Scales. *Journal of School Psychology, 29*, 293-307.
- SPEARMAN, C. (1927). *The abilities of man*. London: Macmillan.
- THURSTONE, L. L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.
- TYLER, S., & ELLIOTT, C. D. (1988). Cognitive profiles of groups of poor readers and dyslexic children on the British Ability Scales. *British Journal of Psychology, 79*, 493-508.

MARK H. DANIEL

DIGIT SPAN See WAIS-R SUBTESTS.

DIGIT SYMBOL See WAIS-R SUBTESTS.

DOLL, EDGAR A. (1889-1968) Edgar Doll was born on May 2, 1889, in Cleveland, Ohio, the son of Arnold Doll and Katherine Rademacher. Doll's education and professional experience combined scientific research on differences in mental ability and personality with a practical concern to apply psychology both to the needs of the mentally retarded and the concerns of American society as a whole. A review of Doll's career will help introduce his most influential contributions to intelligence theory and method of

evaluation—his investigation of adaptive behavior and construction of the Vineland Social Maturity Scale.

Doll graduated from high school in Lakewood, Ohio, and went to Cornell University. There, Guy Whipple, who compiled the first U.S. manual of mental tests, introduced him to psychology. After graduation in 1912, Doll served briefly as an instructor of experimental psychology at the University of Wisconsin before becoming a research and clinical psychologist at the Vineland, New Jersey, Training School. He became a protégé of Henry H. GODDARD, the first director of the Vineland research laboratory and the U.S. translator of the (French) Binet-Simon Intelligence Scale. Doll experimented with a wide variety of mental tests before and during his pursuit of the M.A. in special education at New York University (1916) and the Ph.D. in psychology at Princeton (1920).

After the United States entered World War I in 1917, Doll assisted at the famous Vineland meeting of Robert YERKES's Committee on Psychological Examining; later he served as first lieutenant and psychological examiner. After the war, Doll became chief psychologist and director of the Division of Classification and Education in the New Jersey Department of Institutions and Agencies. This position was similar to one Goddard recently had accepted as director of the Ohio Bureau of Juvenile Research. After Goddard resigned to join the faculty at Ohio State University, Doll became an assistant professor at the same institution in 1923.

Doll returned to Vineland in 1925 to become director of research. He held this position until 1949 (except for 1943 and 1944, when he directed the Bonnie Brae Farm for Boys, in Millington, New Jersey). During this quarter-century, Doll served as president of several professional organizations: the American Association on Mental Deficiency (1935), the American Orthopsychiatric Association (1936), the American Association of Applied Psychology (1940–1941), and the Clinical Division of the American Psychological Association (1945). He was a member of the White House Conference on Child Health and Protection (1931) and, during World War II, the National Research Council's Sub-Committee on Mental Deficiency. He advised numerous state mental health agencies and held visiting and summer teaching positions at several universities in the United States and

Canada. During the last two decades of his life, Doll was research coordinator at the Devereux Schools in Devon, Pennsylvania (1949–1953), and consulting psychologist to the Bellingham, Washington, public schools (1953–1968). Doll was an associate editor of six clinical and educational psychology journals and a prolific writer. In addition to some 250 papers, Doll wrote *Clinical Studies of Feeble-mindedness* (1917), *The Growth of Intelligence* (1920), and *The Measure of Social Competence: A Manual for the Social Maturity Scale* (1953); he was a coauthor of *Mental Deficiency Due to Birth Injuries* (1932) and was editor of the *Handbook of Casework and Classification Methods for Offenders* (1934).

Throughout his writings, Doll defined intelligence after Alfred Binet, as the competence that allowed individuals to learn to adjust and support themselves in adult society, although Binet's work in France was initially based on the testing of mentally retarded children. Before the 1930s, however, Doll adhered to Goddard's early interpretation of Binet's practical concept of "mental (grade) level" as evidence of the recapitulationist evolutionary theory of "mental age." Doll, who first arrived at Vineland as Goddard was publishing *The Kallikak Family*, absorbed his mentor's beliefs that intelligence was inherited and that heredity was responsible for mental deficiency and the social problems that arose from it. Yet Doll also considered nonintellective factors in explaining mental deficiency, since day-to-day familiarity with the "complex symptoms" of many different types of retarded individuals led him to appreciate a plurality of intelligence "kinds" or "types" (Doll, 1914). These appeared across "mental age" levels. When Lewis TERMAN introduced to the United States the intelligence quotient (IQ), or ratio of "mental age" to chronological age to classify students according to a view of intelligence (as academic educability) that was irrelevant to the diagnosis of many of the retarded, Doll (1916) addressed the clinical limitations of Terman's approach. He warned teachers that misuse of the original Binet-Simon Scale had led "amateur psychologists" to ignore the "emotional and volitional aspects of consciousness" (Doll, 1917, p. 6).

Doll's ultimate conversion to a systematic appraisal of noninherited and nonintellective factors did not occur until after 1928, the year in which "mental age" theory clearly failed to explain the new findings of a

Vineland Research Fellow. Ruth Melcher's studies of the birth injured (the cerebral-palsied), a "type" whose mental deficiency origins were most evidently independent of inheritance, inspired development of a remedial muscle-training program. Melcher's innovative use of motion pictures, which documented "everyday acts, (and) . . . reflect(ed) the abilities and disabilities of these children, as well as their improvement under treatment" (Doll, 1932–1933, p. 118), led to new considerations: of the individual as an integral combination of inherited, social, and experiential influences and of the linkage between treatment and growth. The films also encouraged creation of a "schedule of everyday acts . . . as a kind of genetic scale by which all of the birth-injured children (were) scored" (p. 119). What originated as a record of motor progress for individuals with cerebral palsy soon became a "scale of social progress" for use with a wider range of individuals. As Doll later wrote:

We finally concentrated on a different problem, namely, the gains in total performance on the [birth-injured] individual from the point of view of social usefulness, or the practical capitalization of rectified body mechanics as expressed through increased personal adequacy.

Obviously, such a study required anticipation of improvement through growth and development as well as from treatment, and the possibility that these might be interdependent. This led, naturally, to a consideration of maturation versus amelioration. (1953, pp. 4–5)

By 1934, Doll and another research assistant, S. Geraldine Longwell, adapted the "scale of social progress" to "all grades and types of mental deficiency." This happened after it occurred to Doll and Longwell "that the development of such a scale patterned after the general principles of the Binet Scale, and dealing specifically with the measurement of social independence as indicated by social maturity, social competence, and social responsibility, would be an extremely valuable instrument" (1934–1935, pp. 115–116).

The result was the Vineland Social Maturity Scale (VSMS), which advanced a definition of intelligence much closer to Binet's. Echoing his earlier warnings of 1916–1917, Doll committed himself more fully to the belief that "[t]he widespread interest in the measurement of intelligence has overshadowed other phases of human development, and the uncritical emphasis of

IQs has become almost a menace to sound scientific work" (Doll, 1935–1936, p. 2).

The original VSMS of 1935 included a list of 117 items each depicting a different type or level of social performance. The 117 items were formed into an ascending hierarchy of normal behaviors, which most persons would be capable of executing from birth to age 30, in five major categories: General, Eating, Dressing, Locomotion, and Occupation. Because the administration of the 117-item VSMS could be carried out by laypersons as well as psychologists, the scale greatly facilitated supervision of family care. Soon Doll (1936) revised the manual, adding some new performance categories (Self-direction, Communication, and Socialization), and introduced the "SQ" (social quotient), an index (the ratio of "social age" to chronological age) not unlike its IQ counterpart. Within a year, 20,000 copies of two VSMS editions had been distributed (Doll, 1936–1937). Within a decade, the U.S. "standard [for] clinical practice" for the diagnosis of mental retardation entailed administration of both the VSMS and a conventional intelligence test (Matarazzo, 1972, p. 140).

Doll's 1941 definition of "mental retardation" further demonstrated the transition in his thinking, as environmental elements interacted with, but did not eliminate, the inherited ones:

We observe that six criteria by statement or implication have been generally considered essential to an adequate definition and concept [for the diagnosis of mental retardation]. These are (1) social incompetence, (2) due to mental subnormality, (3) which has been developmentally arrested, (4) which obtains at maturity, (5) is of constitutional origin, and (6) is essentially incurable. (Doll, 1941, p. 215)

Before the end of the 1940s, however, Doll withdrew emphasis from the latter three points (Matarazzo, 1972).

As a researcher and applied psychologist throughout his career, Doll saw inadequacies in the theoretical mind-set of psychologists involved in the early professional mental testing movement. Specifically, he examined many individuals who earned identical scores on Terman's IQ test but whose concurrent levels of social maturity and competence varied from very poor to considerably better. This led to his decisive break

with other psychologists around 1930, when he lifted the “social problems” filter from his Goddard-trained vision and looked at individuals—to consider their individual cases, needs and potentials. In so doing, Doll translated Binet’s scientific and practical view of intelligence more effectively than had his mentor, Goddard.

FURTHER READING

A helpful introduction to Doll’s achievements is in an article by his son, E. E. Doll (1962), “A historical survey of research and management of mental retardation in the United States,” in E. P. Trapp and P. Himelstein (Eds.), *Readings on the Exceptional Child* (pp. 21–68), New York: Appleton-Century-Crofts. The best analysis of Doll’s contributions in relation to the dominant strains of intelligence research can be found in Joseph Matarazzo (1972), *Wechsler’s Measurement and Appraisal of Adult Intelligence, 5th and enlarged edition*, New York: Oxford University Press. Although there are no works that specifically address Doll’s transition to a more pluralistic and environmentalist view, the significant change in Goddard’s post-World War I thinking must be considered. For this, see Hamilton Cravens (1987), “Applied science and public policy: The Ohio bureau of juvenile research and the problem of juvenile delinquency, 1913–1930,” in M. M. Sokal (Ed.), *Psychological Testing and American Society: 1890–1930*, New Brunswick, NJ: Rutgers University Press, pp. 158–194. A bibliography of Doll’s most influential publications is in *The Training School Bulletin* (1950), 47, 71–92. More biographical details can be found there and in the unsigned article in the *Cyclopedia of American Biography* (1984) (Vol. 62, p. 162), Clifton, NJ: James T. White.

BIBLIOGRAPHY

- DOLL, E. A. (1914). The kinds of exceptional children. *The Training School Bulletin*, 11, 1–2.
- DOLL, E. A. (1916). Note on the “intelligence quotient.” *The Training School Bulletin*, 13, 36–41.
- DOLL, E. A. (1917). Mental tests and methods of testing. *The Training School Bulletin*, 14, 3–6.
- DOLL, E. A. (1932–33). Department of research annual report, 1931–32. *The Training School Bulletin*, 29, 113–124.
- DOLL, E. A. (1934–35). Department of research annual report, 1933–34. *The Training School Bulletin*, 31, 112–123.

- DOLL, E. A. (1935–36). The Vineland social maturity scale. *The Training School Bulletin*, 32, 1–7, 25–32, 48–55, 68–74.
- DOLL, E. A. (1936). The Vineland social maturity scale: Revised condensed manual of directions. *Publication of the Training School at Vineland, New Jersey, Department of Research*, No. 3.
- DOLL, E. A. (1936–37). Department of research annual report, 1935–36. *The Training School Bulletin*, 33, 118–124.
- DOLL, E. A. (1941). The essentials of an inclusive concept of mental deficiency. *American Journal of Mental Deficiency*, 46, 214–219.
- DOLL, E. A. (1953). *The measurement of social competence: A manual for the Vineland Social Maturity Scale*. Minneapolis, MN: Educational Test Bureau, Educational Publishers.
- MATARAZZO, J. (1972). *Wechsler’s measurement and appraisal of adult intelligence—5th and enlarged ed.* New York: Oxford University Press.

RICHARD T. VON MAYRHAUSER

DOWN SYNDROME Down syndrome (DS) arises from an excess of genetic material located on chromosome 21. An extra chromosome 21, or part of it, produces a cascade of developmental abnormalities, from mental retardation to heart defects to a distinctive facial appearance. DS individuals usually have flattened facial features, small ears, and depressed nasal bridges, often resulting in “epicanthal” folds of skin that partially cover the eye fissure (see Figure 1).

In 95 percent of cases, DS individuals have not only the normal pair of number 21 chromosomes but also an entire, free-floating third copy in their cells. In other cases, all or part of chromosome 21 adheres to another chromosome, often stuck to number 14, 22, or even another 21. Both result in triplicate genetic information, hence the synonym for DS, trisomy 21. In 2 percent of cases, DS is a mosaic of two cell lines, one with trisomy 21 and one with the normal pair of chromosomes (Fryns, 1990).

What part of chromosome 21 is responsible for abnormalities in DS? Mental retardation can result from trisomy of genetic material along the “long arm” of the chromosome 21q (see Figure 2). The critical region responsible for the classical DS facial and cardiac anomalies has been narrowed down to a small area at

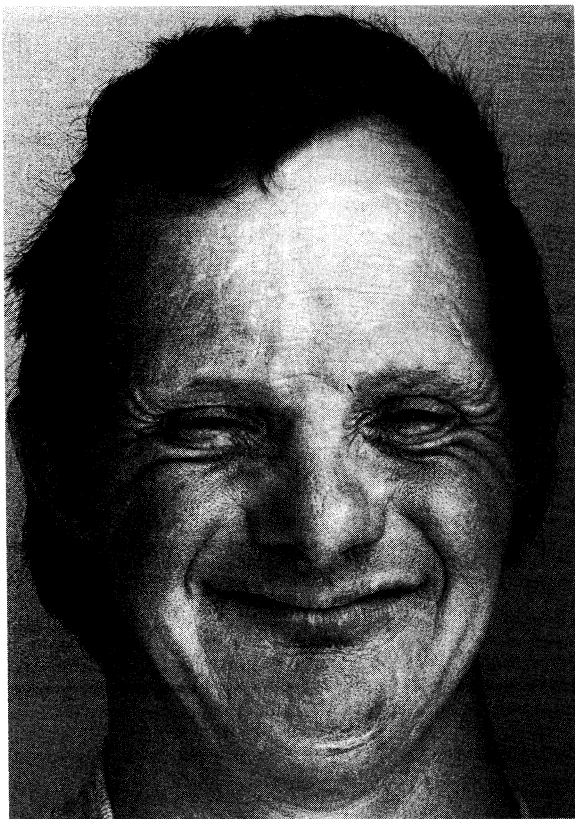


Figure 1
Down syndrome male

the end of 21q (Korenberg et al., 1990). Using the Paris Convention nomenclature (which consecutively numbers the long and short arms of the 23 pairs of chromosomes using distinctive features) the critical DS region spans a proportionately small distance (part of the region from 21q22.1 through 21q22.3).

DS is a common genetic syndrome, occurring in 1 in 770 births. About one-third of DS infants die in their first year, half by school age. Early on, epidemiologists noted that trisomy 21 occurred more often as maternal age increased. DS appears in 1 in 2,500 births to mothers age 20 or younger; but this frequency increases to 1 in 55 in mothers age 45 or older (Fryns, 1990).

A wide range of full-scale IQ scores has been reported for this syndrome. Accurate assessment of intellectual function has been problematic, because conventional IQ tests have not been standardized for

institutionalized mentally retarded populations. They yield marked floor effects, making them useless for severely handicapped individuals (Dalton, 1992). Hearing and speech impediments are common in DS, and if they are uncorrected they may handicap performance on conventional IQ tests (Wisniewski et al., 1988).

In general, DS individuals living in institutions are most frequently moderately to severely mentally retarded, but those raised at home are most frequently mildly to moderately retarded (Wisniewski et al., 1988). This observation and the success of infant stimulation programs have led several investigators to conclude that the DS population includes many individuals with unrealized abilities.

In comparison to normal children, DS individuals do not seem to have smoothly integrated developmental patterns (Hodapp & Zigler, 1988). Rather, the rate of intellectual development slows over time, noticeably after the first year. The reason for this is unknown, but one widely cited hypothesis suggests that retarded language development, which is common in DS, negatively affects further intellectual development (Kopp & McCall, 1982).

Linguistic skills are strongly impaired in this syndrome. Deficits in speech motor control are probably responsible for some of these problems. Diminished muscle tone (hypotonia) produces the tendency for DS individuals to keep their mouths open and protrude their tongues, impeding articulation (Jones, 1988).

Psycholinguistic impairments are conspicuous early. Delayed language production is identifiable by the age of 2. Though there are broad individual differences among DS children, as a group they acquire vocabulary at a much slower rate than would be expected. This problem may lead to a further delay in the start of sentence formation and larger language formations (syntax). Syntactic skills advance even more slowly than vocabulary growth (Miller, 1988).

IQ scores are generally low in DS individuals, but what specific patterns of deficit emerge in this syndrome? Below, separate loci of intelligence are reviewed based on the Sternberg and Spear (1985) triarchic theory of mental retardation. These three phases include (1) information processing and attention; (2) automatization; and (3) environmental adap-

tation. Respectively, 1–3 assess the internal world of the individual (componential loci), the external world of the individual (experiential loci), and the bridge between these worlds (contextual loci).

There is convergent evidence from several lines of research that DS children have significant difficulties in auditory processing relative to visual processing (Pueschel, 1988). Short-term memory is impaired in DS individuals, and auditory sequential (as opposed to simultaneous) processing is particularly poor (Gibson, 1978; Pueschel, 1988). Although visual perceptual skills are comparatively spared in DS, it appears that visual scanning of the environment in order to encode information is worse in these children than in nonretarded children matched for mental age (Hodapp & Zigler, 1990).

There are numerous reports of attentional deficits in DS subjects, often in tasks requiring attention to complex stimuli (Baroff, 1986; Hodapp & Zigler, 1990; Courchesne, 1988; Wagner, Ganiban, & Cicchetti, 1990). Particularly intriguing are numerous reports that DS subjects are *more* visually attentive to simple stimuli than are normal controls. This appears to be true even in infancy—DS babies sometimes appear to be riveted to the stimuli. Electrophysiological studies suggest that old information attracts the same amount of attention as novel, salient information in DS children (Courchesne, 1988). This unusual attentional pattern may well be related to abnormal patterns of habituation (see below).

Most information processing and attention research has been performed on higher-functioning DS participants; stimulus control can be attained more readily with these subjects, and more complex behavioral models can be used. For the same reasons, most studies have been limited to adolescent and adult DS subjects (Wagner et al., 1990). Rigorous stimulus-control methods in combination with new neurobehavioral models may help to delineate DS deficits in lower-IQ and younger subjects (McIlvane, 1992).

Above, we noted that DS individuals sometimes treat old stimuli with the *same* interest given novel ones. Do they also treat the same stimulus as if it were captivating after numerous presentations? Habituation, a type of behavioral automatization after repeated exposure to a stimulus, is heuristic in assessing intelli-

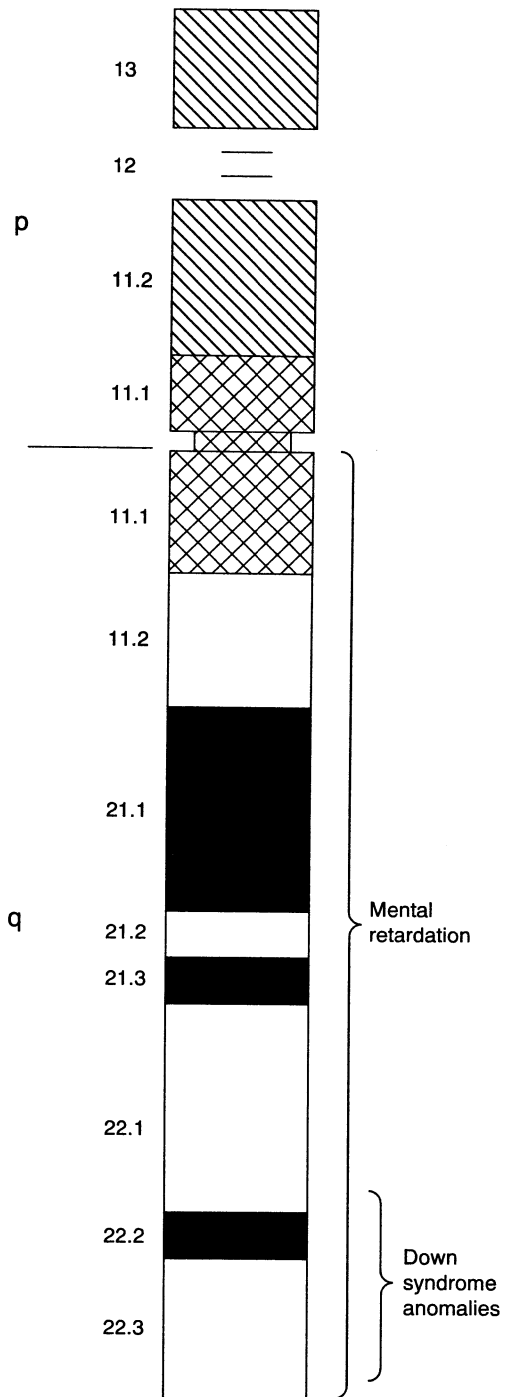


Figure 2
Schematic diagram of the human chromosome number 21
KEY: p = short arm; q = long arm

gence: "It would be hyperbole to say that boredom is the *sine qua non* of intelligence and cognition because to be bored means that one recognizes the present as an instance of the past" (Wagner et al., 1990, p. 152). Several electrophysiological studies do suggest that short-term habituation to simple stimuli is attenuated in DS (Courchesne, 1988), but the interpretation of these studies is controversial (Wagner, 1990).

DS children are relatively strong in social adaptation skills, compared with mental-age-matched controls (Hodapp & Zigler, 1990). Several studies have confirmed the impression of good institutional and community adjustment (Baroff, 1986).

What is the physiological basis of intellectual impairment in this syndrome? There is widespread malformation of the DS brain, involving the cerebral hemispheres, cerebellum, dentate gyrus, and other regions. At the cellular level, there is prenatal retardation of the development of nerves and synapses; from birth, DS brains show fewer neurons and lower neuronal densities. Causal links between regional brain pathology and specific behavioral abnormalities are still speculative (Courchesne, 1988).

The brain abnormalities described above result from early maldevelopment. By the time the DS individual reaches middle age, an *additional* form of neuropathology arises—one that is almost indistinguishable from pathology found in the brains of Alzheimer's disease patients. Most, perhaps all, DS individuals over age 35 years have Alzheimer's-like neuropathology. Despite the prevalence of this neuropathology, only approximately 40 percent of these individuals show frank clinical signs of dementia (Schapiro et al., 1992). Perhaps a more rigorous behavioral assessment will reveal a higher rate of dementia in DS (Dalton, 1992).

BIBLIOGRAPHY

- BAROFF, G. F. (1986). *Mental retardation: Nature, cause, and management* (2nd ed.). New York: Hemisphere Publishing.
- COURCHESNE, E. (1988). Physioanatomical considerations in Down syndrome. In L. Nadel (Ed.), *The psychobiology of Down syndrome* (pp. 291–313). Cambridge, MA: MIT Press.
- COYLE, J. T., OSTER-GRANITE, M. L., & GEARHART, J. D. (1986). The neurobiologic consequences of Down syndrome. *Brain Research Bulletin*, 16, 773–787.
- DALTON, A. J. (1982). Dementia in Down syndrome: Methods of evaluation. In L. Nadel & C. J. Epstein (Eds.), *Down syndrome and Alzheimer disease* (pp. 51–76). New York: Wiley-Liss.
- FRYNS, J.-P. (1990). Down syndrome, trisomy 21. In M. L. Buyse (Ed.), *Birth defects encyclopedia* (pp. 391–393). Cambridge, MA: Blackwell Scientific Press.
- GIBSON, D. (1978). *Down's syndrome: The psychology of mongolism*. New York: Cambridge University Press.
- HODAPP, R. M., & ZIGLER, E. (1990). Applying the developmental perspective to individuals with Down syndrome. In D. Cicchetti & M. Beeghly (Eds.), *Children with Down syndrome: A developmental perspective* (pp. 1–28). New York: Cambridge University Press.
- JONES, K. L. (1988). Down syndrome (trisomy 21 syndrome). In K. L. Jones (Ed.), *Smith's recognizable patterns of human malformation* (4th ed.). Philadelphia: W. B. Saunders.
- KOPP, C., & MCCALL, R. (1982). Predicting later mental performance for normal, at risk, and handicapped infants. In P. Baltes & O. Brim (Eds.), *Lifespan development and behavior 4*, New York: Academic Press.
- KORENBERG, J. R., KAWASHIMA, H., PULST, S. M., ALLEN, L., MAGENIS, E., & EPSTEIN, C. J. (1990). Down syndrome: Toward a molecular definition of the phenotype. *American Journal of Medical Genetics* (Suppl.), 7, 91–97.
- MATTHYSSE, S., & DEUTSCH, C. K. (1991). Convergence and chaos in morphogenesis. In C. J. Epstein (Ed.), *The morphogenesis of Down syndrome. Progress in Clinical and Biological Research*, 373, 19–41.
- MCILVANE, W. J. (1992). Stimulus control analysis and non-verbal instructional methods for people with intellectual disabilities. In N. R. Bray (Ed.), *International Review of Research in Mental Retardation*, 18, New York: Academic Press, 55–109.
- MERVIS, C. B. (1990). Early conceptual development of children with Down syndrome. In D. Cicchetti & M. Beeghly (Eds.), *Children with Down syndrome: A developmental perspective*. New York: Cambridge University Press.
- MILLER, J. F. (1988). The developmental asynchrony of language development in children with Down syndrome. In L. Nadel (Ed.), *The psychobiology of Down syndrome*. Cambridge, MA: MIT Press.
- PUESCHEL, S. M. (1988). Visual and auditory processing in children with Down syndrome. In L. Nadel (Ed.), *The psychobiology of Down syndrome* (pp. 199–216). Cambridge, MA: MIT Press.

- SCHAPIRO, M. B., HAXBY, J. V., & GRADY, C. L. (1992). Nature of mental retardation and dementia in Down syndrome: Study with PET, CT, and neuropsychology. *Neurobiology of Aging*, 13, 723-734.
- STERNBERG, R. J., & SPEAR, L. C. (1985). A triarchic theory of mental retardation. *International Review of Research in Mental Retardation*, 13, 301-326.
- WAGNER, S., GANIBAN, J., & CICCETTI, D. (1990). Attention, memory, and perception in infants with Down syndrome: A review and commentary. In D. Cicchetti & M. Beeghly (Eds.), *Children with Down syndrome: A developmental perspective* (pp. 147-179). New York: Cambridge University Press.
- WISNIEWSKI, K. E., MIEZEJESKI, C. M., & HILL, A. L. (1988). Neurological and psychological status of individuals with Down syndrome. In L. Nadel (Ed.), *The psychobiology of Down syndrome*. Cambridge, MA: MIT Press.

CURTIS K. DEUTSCH

DRAW-A-FIGURE TEST Since the late 1800s, professionals interested in assessing intelligence have been intrigued by the correspondence between children's development and the increasing complexity of their drawings. The possibility that children's cognitive abilities could be estimated from their drawings led to the use of *human figure drawings* (HFDs) as standardized *intelligence tests*. Later, HFDs were also widely accepted as a projective method of emotional and personality assessment (Lubin, Larsen, & Matarazzo, 1984).

Many qualities of HFD tests, such as their non-threatening and enjoyable characteristics and their simple scoring procedures, have resulted in their popularity for the assessment of intellectual development, primarily for 3- to 10-year-olds. HFD tests are widely recommended for use as brief *cognitive or developmental screening measures* to identify the need for more comprehensive assessment. Professionals routinely incorporate HFD tests into multidimensional assessment batteries for children experiencing developmental or academic problems. Because HFD tests require no reading or verbal responses, they are often used as measures of *nonverbal intelligence* for children with hearing impairments or language disabilities. HFD tests may be influenced less by culture, language, and academic experience than are traditional intellectual mea-

asures, and they are often recommended for assessment of children with diverse backgrounds. HFD tests are occasionally used to assess the abilities of adolescents and adults with mental retardation or neurological impairments.

VERSIONS

The appealing and enduring qualities of HFD tests have resulted in the development of a number of different versions over the past seventy years. Goodenough (1926) developed the first standardized approach to analyzing children's HFDs and reported that intellectual development was a chief factor in the quality of children's drawings. Her test, known as the Draw-A-Man Test, was a psychometric feat for its day (Kamphaus & Pleiss, 1991). Harris (1963) expanded and revised Goodenough's test. His Goodenough-Harris Draw-A-Person Test added several features to the older system, including more extensive and objective scoring, drawings of a woman and self, national norms, and replacement of Goodenough's mental age IQ with a deviation IQ and standard scores. Koppitz's (1968) HFD developmental scoring system incorporated many of the scoring criteria from the Goodenough-Harris version. She also offered emotional indicators for projective interpretation.

The success of HFD tests led to their inclusion in a number of developmental screening batteries and intelligence tests for young children. For example, the Gesell School Readiness Test, the Denver Developmental Screening Test, and the McCarthy Scales of Children's Ability all require children to draw human figures. Two new HFD tests were recently developed in the tradition of the Goodenough-Harris Draw-A-Person Test: the Human Figures Drawing Test (Gonzales, 1986), and the Draw-A-Person: A Quantitative Scoring System (Naglieri, 1988).

ADMINISTRATION AND SCORING

Administration and scoring for HFD tests are relatively simple. For all versions, general procedures for estimating intelligence are similar. Children are usually asked to draw a man, woman, child, or self-portrait. There are no time limits. After drawings are completed, examiners are allowed to ask questions in order to clarify ambiguities. Drawings are then scored ac-

cording to detailed quantitative systems that assign points for inclusion of specific features (e.g., body parts, facial features, clothing, etc.). In most versions, total points are compared to points earned by a normative sample of other children the same age and are transformed to standard scores with a mean of 100 and a standard deviation of 15.

RELIABILITY AND VALIDITY

Much concern and debate surround the psychometric qualities of the HFD tests when they are scored for intelligence. RELIABILITY estimates for most versions are reported to be in the adequate range, considering the brevity of the tests. Test-retest and internal-consistency coefficients are generally in the .60s to .80s. Interrater reliability coefficients are somewhat higher, with most studies reporting correlations of .80s to .90s between different scores.

The VALIDITY of HFD tests has been investigated in numerous studies. Many studies have reported high intercorrelations between different versions of HFD tests. However, results of validity studies have not supported the use of HFDs as measures of intelligence or development or as predictors of academic performance. Correlations between HFD tests and comprehensive intelligence tests such as the Wechsler and Stanford-Binet scales are generally in the low-to-moderate range. Correlations between HFD tests and developmental measures, such as the Matching Familiar Figures Test and the Peabody Picture Vocabulary Test, are also in the low-to-moderate range. Correlations between HFD tests and academic performance tend to be low. Research has not found that HFD tests effectively identify children who experience academic problems or who have developmental disabilities. Furthermore, there is no evidence to support the treatment validity of HFD tests; HFD tests do not appear to facilitate effective decisions about children with learning problems or disabilities, nor do they offer information for planning effective interventions.

CONCLUSIONS

The efficiency of HFD tests has resulted in their popularity for many years; thus, HFD tests will probably continue to be a part of many assessment batteries in the future (Cosden, 1992). Unfortunately, studies

suggest that most of the recommended uses of HFD tests are not valid. The lack of a strong relationship between HFD tests and important intellectual and academic criteria indicate that HFD tests may not be valid for use as screening measures, brief measures of intelligence or development, or predictors of school performance. In spite of the appeal of HFD tests, they may have little utility in measuring children's intelligence.

BIBLIOGRAPHY

- COSDEN, M. (1992). Review of the Draw-A-Person: A Quantitative Scoring System. In J. J. Kramer & J. C. Conoley (Eds.), *The eleventh mental measurement yearbook* (pp. 287-289). Lincoln, NE: The Buros Institute of Mental Measurements.
- GONZALES, E. (1986). *Human figures drawing test*. Austin, TX: PRO-ED.
- GOODENOUGH, F. L. (1926). *Measurement of intelligence by drawings*. New York: World Book.
- HARRIS, D. B. (1963). *Children's drawings as measures of intellectual maturity*. New York: Harcourt, Brace & World, Inc.
- KAMPHAUS, R. W., & PLEISS, K. L. (1991). Draw-A-Person techniques: Tests in search of a construct. *Journal of School Psychology, 29*, 395-401.
- KOPPITZ, E. M. (1968). *Psychological evaluation of children's human figure drawings*. New York: Grune and Stratton.
- LUBIN, B., LARSEN, R. M., & MATARAZZO, J. D. (1984). Patterns of psychological test usage in the United States: 1935-1982. *American Psychologist, 39*, 451-453.
- MCCARTHY, D. (1972). *McCarthy scales of children's ability*. New York: The Psychological Corporation.
- NAGLIERI, J. A. (1988). *Draw A person: A quantitative scoring system*. San Antonio, TX: The Psychological Corporation.
- OSTER, G. D., & GOULD, P. (1987). *Using drawings in assessment therapy: A guide for mental health professionals*. New York: Brunner/Mazel.

PATTI L. HARRISON
HANNELORE H. SCHOCK

DRUGS AND INTELLIGENCE Psychologists define intelligence in many ways. The primary characteristics of intelligence are the abilities to learn from experience, to respond quickly and successfully to new situations or novel stimuli, and to solve prob-

lems. Intelligence may also be defined as the highest level of neural integration in the brain as expressed by the action of knowing, perceiving and conceiving, and as opposed to emotion and volition. Each of these characteristics of intelligence can be influenced by drugs that alter specific aspects of brain function. The interrelationship between drugs and brain function, as expressed by behavior, delineates a subject of current experimental and clinical interest. Scientists refer to these drugs as "psychotropic," that is, acting upon the brain to affect psychological function. Psychotropic drugs can either impair or enhance the ability of an individual to demonstrate one or many of the characteristics of intelligence.

Scientists interested in the neural mechanisms underlying intelligence usually conduct experiments upon laboratory animals, typically rats and monkeys. Laboratory animals demonstrate many of the same characteristics of intelligence as humans and can provide useful information on the mechanisms of action and effects of specific drugs. The results of these studies are summarized below.

One important development has been the solution to the mystery of the anatomy of the brain and the recognition of the role of its individual components and neural systems, particularly the neurotransmitter pathways, in intelligence. Neural pathways form the communication lines between different brain regions. At their points of communication with each other nearly all of the important drug-brain interactions occur; these points of communication are called the *neuronal synapses*. Within these synapses are specific proteins that act as recognition sites, or receptors, for the different drugs. The brain has many such pathways that communicate by releasing specific chemicals, called *neurotransmitters*. These neurotransmitters are made by the brain from nutrients in the diet. Indeed, the balance of these dietary nutrients or their imbalance can influence brain function and intelligence dramatically by their ability to alter the concentration of the neurotransmitter molecules. For example, increased intake of the amino acid tryptophan (once available in health food stores) is associated with a slight increase in the production of neurotransmitter serotonin within the brain. Increased production of serotonin can transform the brain from a state of alertness to one of drowsiness or sleep. Serotonin is also

important for controlling mood. Some recreational drugs, such as amphetamine, can relieve temporarily some symptoms of depression because they enhance the release of serotonin, as well as many other neurotransmitters. The psychedelic drugs LSD, mescaline, and psilocybin induce hallucinations by way of their actions on neurons that produce serotonin.

HALLUCINOGENS

Little reliable information is available about the effects of the hallucinogens, such as PCP (phencyclidine piperidine, or angel dust), bufotenin, DMT (*n,n*-dimethyltryptamine), ecstasy (3,4-methylenedioxyamphetamine), LSD (lysergic acid diethylamide), mescaline (from peyote—the cactus *Lophophora williamsii*) or psilocybin (from the mushroom *Psilocybe mexicana*), upon intelligence (Judd et al., 1987). In most cases, the actions of these drugs upon the brain are so profound that normal intellectual processes are significantly impaired until the drugs leave the brain and/or the body. Indeed, some drugs, such as LSD, continue to produce hallucinations long after they are no longer detectable within the brain. For these reasons, the determination of the effects of hallucinogens upon intelligence is very difficult. The consequences of long-term use of most hallucinogens has never been completely determined. Scientists believe that prolonged and daily PCP use can produce short-term memory deficits; in addition some individuals have reported significant difficulties with speech up to one year after cessation of regular use. The pathological effects of the recreational drug ecstasy upon the brain are fairly well understood; its long-term effects upon specific cognitive functions that underlie intelligence are not. In both rodents and nonhuman primates, a single high dose or a few low doses (approximating those taken by humans) of ecstasy irreversibly destroys the population of neurons that produce serotonin. Numerous attempts have been made in the laboratory to understand the effects of ecstasy upon some aspect of intelligence. However, rodents and nonhuman primates given ecstasy or humans who have used ecstasy, do not demonstrate a significant change in learning and memory abilities, mood, or sleeping habits, that is, those brain functions that may require serotonin neurons. The consequences of ecstasy use may be appreciated

only as these subjects become older and their brains' abilities to compensate for past injuries diminishes.

Long-term, daily marijuana use impairs short-term memory or recent memory, particularly those events that occur during or immediately after its use. Researchers have determined that the ability of individuals to retrieve information that is already in long-term memory, for example, memory of the rules of language or arithmetic, is not altered by marijuana intoxication. Nonetheless, marijuana intoxication impairs the consolidation process, that is, the transformation of short-term memories to long-term storage. Some users often compare the effects of marijuana to those of alcohol. However, the effects of marijuana upon brain function are quite different from those of alcohol intoxication. In contrast to marijuana, which does not depress central nervous system functioning, alcohol does have this effect and therefore prevents normal information processing of any kind from occurring. Alcohol makes walking a straight line difficult; marijuana does not. Both alcohol and marijuana intoxication can prevent the user from engaging in a complicated mental or verbal task. The impairment in short-term memory produced by marijuana may also underlie changes in time-sense that has been reported by recreational users. The user feels that time is accelerated. According to some reports, marijuana also increases sensory awareness or sensitivity to common sensory stimuli. However, recent scientific investigations have not been able to confirm this claim. The marijuana user's response to sensory input may be altered, but the ability of the brain to receive and analyze sensory information through normal physiological mechanisms does not appear to be altered.

STIMULANTS

Drugs that stimulate the brain tend to enhance performance in tasks that test attention, an important aspect of intelligence. The best studied stimulants are caffeine, amphetamine, cocaine, and nicotine.

Caffeine has many beneficial effects. Primarily it enhances mental clarity and allays fatigue. Caffeine is most effective in improving performance that has deteriorated because of excessive stress or fatigue. It has a much less dramatic benefit on well-rested individuals. Interestingly, caffeine has a more pronounced ben-

efit upon the performance abilities of highly impulsive people as compared to less impulsive people. It also increases vigilance and prevents the decline in attentional ability that is frequently seen after meals; in particular, it can significantly improve information processing after lunch. Caffeine also enhances the ability of subjects to respond to two different stimuli. The performance of women engaged in intelligence tasks are enhanced most by low doses of caffeine in the first five days of their menstrual cycle suggesting a potential interaction with the body's hormones. In all cases studied, caffeine does not actually improve intellectual abilities; rather it enhances an individual's ability to focus attention and thereby improves performance of difficult tasks.

The beneficial effects of caffeine vary according to the age of the subject studied. Within a few hours after taking a relatively high dose (about 2-3 cups of coffee) of caffeine, elderly subjects made fewer errors on an attention task and had a faster response rate in a reaction-time task. When given the same dose, young adults felt more alert, calmer, and more interested in performing a complicated or difficult task. A low dose (a few sips of coffee) of caffeine caused children to speak faster, have a faster reaction speed, and make fewer mistakes on attention tasks. In general, however, elderly subjects tend to show a stronger response to caffeine than do younger subjects.

Amphetamine enhances performance in many different behavioral tasks that require learning and memory or increased vigilance (Judd et al., 1987). The brain's utilization of its primary energy source, glucose, is greatly increased by amphetamine. In addition, the electrical activity measured at the scalp, the EEG, is enhanced. This enhancement correlates with improved performance in a variety of tasks that measure learning and memory. Amphetamine has similar cognitive effects in both normal children and those with ATTENTION-DEFICIT HYPERACTIVITY DISORDER. Both show improvements in tests of recall of previously learned information, arithmetic abilities, and vigilance. The ability of amphetamine to enhance cognitive function may in fact be related to its actions outside the central nervous system. Research on animals has shown that peripheral administration of amphetamine (taking a pill or receiving an injection) enhances memory; in contrast, injection of amphetamine directly into the

brain does not. The effects of amphetamine on intelligence may derive from its ability to increase blood glucose levels quickly. These peripheral actions of amphetamine are completely unrelated to its actions within the brain that lead to excitation and euphoria.

Cocaine is also a powerful brain stimulant that enhances performance of laboratory rats in selected behavioral tasks requiring learning and memory or vigilance. Few well-controlled studies have been conducted on humans. In one study, cocaine disrupted learning; in another, it increased the reaction time of subjects who were sleep-deprived but not of those who were well rested. The similarities between the actions of cocaine and amphetamine in the brain may underlie their similar effects on intelligence. A warning is in order. In these experimental studies very low doses of amphetamine and cocaine were given only once or twice to naive rats—rats that had never experienced either drug. (These doses were well below the recreational doses typically taken by humans.) Repeated exposure to high doses of either drug actually impaired performance on these same behavior tasks. This situation may result from compensatory changes that the brain makes in response to the continued presence of the drug. These compensatory changes are not well understood but may develop according to the same rules that govern learning in the brain, such as specific changes in the ways that individual neurons communicate with each other.

Nicotine is also a powerful central nervous system stimulant. Numerous studies on humans have suggested that nicotine may enhance short-term memory. Other research has refined these early interpretations. These investigations suggest that nicotine influences overall intelligence by increasing a subject's speed of response in selected tasks, enhancing their ability to focus quickly upon relevant visual information and by improving overall attention and information processing rather than by enhancing any particular memory process within the brain (Sherwood, Kerr, & Hindmarch, 1990).

DEPRESSANTS

In contrast to these stimulants, drugs that depress the function of the brain, such as opiates (heroin and morphine), alcohol, barbiturates and anxiolytics (drugs

that reduce anxiety, such as Valium, Librium, and related drugs) tend to impair performance on intelligence tasks.

The scientific literature on the opiates is very confusing. In laboratory studies, opiates tend to interfere with learning and memory, and drugs that block the actions of opiates in the brain actually may improve learning and memory abilities and/or enhance attention. In some studies, higher cognitive functions were not affected by low doses of opiates but were impaired at high doses. The timing of the administration of the opiate is also important—whether the opiate was taken before or after the subject attempted to perform an attention task or learn new information. Studies have also shown that opiate antagonists (drugs that prevent opiates from working within the brain) can produce mental slowing, although intellectual abilities do not diminish and may even increase in some circumstances.

Alcohol and the barbiturates depress the activity of neurons within the brain. They produce such profound changes in brain function that the determination of intellectual abilities is very difficult. Both types of drugs tend to release behavior that had been suppressed previously by punishment. Assuming that these drugs are not used repeatedly, intellectual ability after their use usually does not decline. Nonetheless, disorders of memory and critical thought processes have been associated with chronic use.

The effects of drugs like Valium and Librium upon specific aspects of cognitive function are easier to investigate than the actions of most other central nervous system depressants. Valium, Librium, and their related drugs, have a similar major side effect: drowsiness. Their primary effects upon intellectual ability probably relate to this drowsiness. Studies have found a diminution of specific attentional abilities, such as tracking eye-movements (the ability of an individual to follow the movements of an object in the environment with the eyes). For example, tracking eye-movements are important for driving a car. Alcohol also causes a tracking impairment. A typical dose of Valium (2–10 mg) produces significant impairments in driving ability—lane-tracking, lane-changing, and stopping—for up to 3.5 hours after taking the drug. Valium, Librium, and similar anxiolytic drugs may also act in a specific brain region such as the hippocampus, which is im-

portant for learning and memory, and may produce temporary amnesia. However, studies suggest that this amnesia may be related more closely to the drowsiness produced by these drugs than to their action in the hippocampus.

DRUGS THAT IMPROVE INTELLIGENCE

Normal individuals do not always perform perfectly or even maximally. Some drugs, called *cognitive enhancers*, may actually be able to improve normal brain function and imperfect performance. Many drugs have been tested for their ability to enhance the cognitive processes that underlie intelligence in laboratory animals and humans. Enhancement of the function of neurons that produce acetylcholine has received most emphasis because of its demonstrated role in the neural processes that underlie learning and memory. Drugs that impair the function of these neurons, such as extracts of plants from the genus *Datura* or *Atropa*, also impair performance of humans and animals in tasks that require learning and memory, two important components of intelligence. In contrast, drugs that enhance the function of acetylcholine-containing neurons, such as over-the-counter preparations of choline and lecithin (commonly found in health food stores) tend to improve the performance of laboratory animals in these tasks. Many of the drugs that enhance the function of acetylcholine-containing neurons have been tested on patients with Alzheimer's disease, a disorder characterized by a significant loss of acetylcholine neurons, but with very limited success (Iversen, 1985).

Another class of intelligence-enhancing drugs, *nootropics*, such as piracetam, anaracetam, or oxiracetam, are often one of the many ingredients combined into beverages that are sold as "Think Drinks" in "smart bars." These drinks usually contain a few amino acids, choline, caffeine, and a pharmaceutical agent such as piracetam, deprenyl (which inhibits the breakdown of the brain's neurotransmitters), Dilantin (used to treat epilepsy), or Hydergine (once used to treat high blood pressure). Some drinks may also contain antioxidants, such as vitamin E or Idebenone, a Japanese pharmaceutical product. Without exception, the only people who benefit from the consumption of these drinks are

the people who sell them. The mixtures themselves are usually harmless. Excess consumption of some of the ingredients, however, combined with the use of certain nonprescription or prescription drugs, can have serious consequences. Very little is known about both the mechanisms of actions of many of these drugs and their interactions with the thousands of other drugs that people take daily.

DRUGS THAT INFLUENCE CREATIVITY

The information obtained for many of these drugs derives from examination of their effects on lower mammals. The determination of their effects on humans requires a generalization from the study of the lower mammals to our species. This extrapolation complicates the interpretation of the experimental findings in many areas of research on cognition and intelligence. Studies of the effects of drugs on creativity, an important expression of intelligence, do not suffer from problems of generalization to our species; most of the artistic products that have aesthetic value or interest to humans are works of humans. The real limitation on the study of the effects of recreational drugs on creativity, as an expression of intelligence, has been the legal restrictions. A serious lack of criteria for evaluating creative efforts also plagues this field of research. The relative merits of particular drugs to enhance creativity are usually in direct proportion to the personal interest of the viewer in the continued use of the drug.

Some drugs may enhance creativity, not by their ability to enhance some unknown mental process but by their ability to release the user from the constraints of another problem, either mental or physical. A sample situation is the relief from physical or mental pain, anguish, anxiety, depression, or severe personal problems. An excellent example of a recreational drug used in this way might be alcohol, which may have released the creative genius of many artists, such as Thomas Wolfe, Dylan Thomas, and F. Scott Fitzgerald, from physical and emotional pain and anxiety. The popular prescription drugs valium and librium, and their various analogues, have provided immeasurable relief from anxiety for many other artists.

The more familiar recreational stimulants, such as caffeine and chocolate, as well as the more potent am-

phetamine, have probably sustained many persons during periods of fatigue to complete a specific creative task. These agents are most effective for single instances of enhanced effort and can have unpleasant consequences when used repeatedly. For example, excessive caffeine can lead to nervous irritability and headaches during withdrawal. Chocolate is also addicting and can, like caffeine, exacerbate the irritation associated with fibrocystic disease of the breast, in addition to increasing cholesterol levels in the blood. Animal research has clearly shown that extended amphetamine use leads to a repetitive behavioral pattern similar to that seen in psychotic patients. This behavior would certainly be incompatible with the creative state.

The psychotropic drugs that have been used most often to enhance creativity are LSD, psilocybin, the amphetamine-like agent DMT, morning glory seeds (containing the LSD-like drug lysergic acid monoethylamide), the *Amanita muscaria* mushroom, and peyote (Leavitt, 1982). These recreational drugs generate an altered state of consciousness that the users claim may lead to enhanced creative abilities.

Many side effects of the drugs may actually hinder the creative experience. For example, the anxiety and physical agitation produced by many stimulants, such as amphetamine or many psychedelics, among them mescaline and LSD, either may prevent the user from recalling the creative experience or simply may not allow the user to benefit from the drug-induced experience. Sometimes the users simply become too self-absorbed in the drug-induced experience to appreciate the aesthetic qualities of the "trip."

Most of the results from studies on the ability of recreational drugs to improve creativity have been negative. Nonetheless, it is difficult to make conclusive statements about negative outcomes of studies that investigate creativity. Scientists do not know whether the tests are sensitive to the effects of the drugs or specific for creativity and not for some other measure of intelligence. The scientific literature has provided little insight into how drugs such as mescaline, marijuana, and LSD have helped various artists, dancers, and architects to produce creative and innovative products. For example, some experimenters or subjects have profound biases about the presumed outcome of the investigations. The few accounts of

famous artists who have benefited from these drugs probably became more noteworthy than the many other drug-related experiences that did not generate important artistic products. Recreational drug use may not enhance creativity, but, it may not be incompatible with the creative process either. The experiences of authors who have claimed benefit from recreational drug use, such as Alan Watts (*As a Man Thinketh*), Ken Kesey (*One Flew over the Cuckoo's Nest*), and William Burroughs (*Naked Lunch*), may demonstrate only that individuals who choose to live an unconventional life may also choose unconventional experiences like the altered mental state produced by these drugs (Masters & Houston, 1968).

CONCLUSIONS

Any drug that influences brain function must have an effect on some aspect of intelligence. The nature of this effect depends on the particular neurotransmitter system that is effected within the brain. Short-term use of drugs at low doses tends to produce much smaller effects than long-term use at higher doses. The effects of some drugs are so profound that any attempt to measure intelligence is impossible. Clearly then, although some drugs may temporarily improve performance, most drugs tend to impair, rather than enhance, overall intelligence when they act on the brain.

Whether drugs can truly enhance attention, learning and memory, creativity, or any other aspect of intelligence is currently unknown. There are two potential problems associated with trying to enhance intelligence by altering normal brain function. First, the brain may already be functioning at its maximal level of performance. It simply may not be possible to improve upon millions of years of evolution by the administration of a single chemical to such a complicated nervous system. Second, the individual neural systems that comprise the brain function in a subtle balance. This balance is maintained by millions of very brief interactions occurring throughout millions of neural networks. The gross manipulation of any one of these neural systems by a recreational drug usually imbalances the system, enhancing a single measure of intelligence but impairing severely other cognitive functions that contribute to one's overall intelligence.

BIBLIOGRAPHY

- GOODMAN AND GILMAN'S: THE PHARMACOLOGICAL BASIS OF THERAPEUTICS, (8th ed.). (1990). New York: Pergamon Press. An excellent resource book on all aspects of drug effects upon the body.
- IVERSEN, S. D. (1985). Psychopharmacology of cognition. In S. D. Iversen (Ed.), *Psychopharmacology: Recent advances and future prospects*. New York: Oxford University Press.
- JUDD, L. L., SQUIRE, L. R., BUTTERS, N., SALMON, D. P., & PALLER, K. A. (1987). Effects of psychotropic drugs on cognition and memory in normal humans and animals. In H. Y. Meltzer (Ed.), *Psychopharmacology: The third generation of progress*. New York: Raven Press. This book is an excellent reference source for many aspects of psychopharmacology.
- LEAVITT, F. (1982). *Drugs and behavior* (2nd ed.). New York: Wiley.
- MASTERS, R., & HOUSTON, J. (1968). *Psychedelic art*. New York: Grove.
- SHERWOOD, N., KERR, J. S., & HINDMARCH, I. (1990). The effects of nicotine on human memory. *Journal of Drug Development*, 3(3) 121-126.
- WENK, G. L., & OLTON, D. S. (1989). Cognitive enhancers: potential strategies and experimental results. *Progress in neuro-psychopharmacology & biological psychiatry*, 13 (Suppl.), S117-S139.

GARY L. WENK

DYNAMIC ASSESSMENT OF MENTAL ABILITIES

Since the late 1970s the dynamic assessment of mental abilities has drawn the attention of many psychologists and educators who have become dissatisfied with intelligence quotient (IQ) tests. Dissatisfaction with IQ tests arises primarily because such measures fail to provide useful information about children's learning potential or about cognitive and meta-cognitive processes. In addition, the classification of children into ability groups based on IQ assessment is of questionable utility, since little guidance is provided for the practitioner responsible for the development and implementation of instructional programs. By providing insights into how individuals learn and the extent to which cognitive processes can be modified by interventions, dynamic assessment represents an im-

portant alternative to traditional methods of assessment.

Dynamic assessment, often called interactive assessment, is not a singular measurement technique of mental abilities or a theoretical approach but rather encompasses a range of methods of assessment that share a number of characteristics. Some common characteristics are the following:

1. The performance of the individual being tested can be modified through the intervention of the tester.
2. Cognitive competence is distinguished from cognitive performance.
3. Enhancing test performance through experimenter intervention and/or modifications in testing procedures yields better estimates of cognitive competence than do traditional testing methods.
4. The processes of thinking and problem solving are considered to be at least as important as their products.
5. Assessment of abilities that are in the process of developing facilitates accurate predictions about future development.

VYGOTSKY'S INFLUENCE ON DYNAMIC ASSESSMENT

Perhaps the most significant individual contribution to the dynamic assessment movement comes from the writings of the early-twentieth-century psychologist Lev Semenovich VYGOTSKY. Although appreciative of the experimental approach and the necessity of using standardized methods of assessment, Vygotsky criticized traditional testing procedures, arguing that it is more important to investigate the *processes* of knowledge acquisition than the *products* of previous learning. For Vygotsky, the interesting questions about cognitive assessment relate to determining one's learning potential and how one acquires knowledge. These considerations are based on a notion fundamental to Vygotsky's psychological theory: the zone of proximal development.

The Zone of Proximal Development. According to Vygotsky, the zone of actual development is what children can do on their own, that is, without the aid of another person. The zone of proximal development is the discrepancy between what children

can do unaided and what they can do with assistance. Accurate assessment requires that attention be paid to what children can accomplish through interaction with more competent peers and/or adults. If two children score the same on an IQ test, it does not mean that their zones of proximal development are the same. With assistance, Vygotsky maintains, one child may attain a level of performance far surpassing that of the other child. It is the level of performance after intervention that the educator must consider. Vygotsky wrote:

Suppose I investigate two children upon entrance into school, both of whom are ten years old chronologically and eight years old in terms of mental development. Can I say that they are the same age mentally? Of course. What does this mean? It means that they can independently deal with tasks up to the degree of difficulty that has been standardized for the eight-year-old level. If I stop at this point, people would imagine that the subsequent course of development and of school learning for these children will be the same, because it depends on their intellect. . . . Now imagine that I do not terminate my study at this point, but only begin it. . . . Suppose I show . . . [these children] various ways of dealing with the problem . . . that the children solve the problem with my assistance. Under these circumstances it turns out that the first child can deal with problems up to a twelve-year-old's level, the second up to a nine-year-old's. Now are these children mentally the same?

When it was first shown that the capability of children with equal levels of mental development to learn under a teacher's guidance varied to a high degree, it became apparent that those children were not mentally the same age and that the subsequent course of their learning would obviously be different. This difference between the twelve and eight, or between nine and eight, is what we call the zone of proximal development. It is the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers. (Vygotsky, 1935/1978, pp. 85–86)

Vygotsky's notion of the zone of proximal development is consistent with learning theories that as-

sume that test reliability is increased by allowing the subject to practice and overlearn the task material. The reason for this is simply that for a measure to be reliable it must yield consistent results, and consistency is enhanced by employing methods of assessment that maximize performance.

MODELS OF DYNAMIC ASSESSMENT

Several models of dynamic assessment have been developed. Although there is some variation in the extent to which these models adhere to Vygotsky's theory, all are consistent with the main tenets of his thought.

Budoff's Learning Potential Assessment.

Concerned that traditional IQ assessment leads to inappropriate special education placement for some children, Milton Budoff and his colleagues developed an alternative strategy for assessing cognitive functioning, which they termed "learning potential assessment" (see Budoff, 1987). Learning potential assessment is a training-based paradigm in which the child is first tested in the traditional manner and then given training to improve competence with the materials. The child is then retested.

Training involves familiarizing the subject with the tasks and their demands and appropriate solution strategies, with a view to increasing the child's confidence. The instruments Budoff and his colleagues use include traditional measures such as the Raven Progressive Matrices and Kohs Blocks, although during the training phase the method of administering these measures is very different from standard modes of presentation.

The test-train-retest format allows Budoff to distinguish between *gainers* and *nongainers*. Gainers are children whose posttest score exceeds their pretest score. The improved posttest performance is seen to represent the child's optimal level of performance and provides the teacher or curriculum planner with information about the child's learning potential. Nongainers are children for whom the initial score seems to be a valid representation of their abilities.

Campione and Brown's Guided Learning and Transfer Model. Joe Campione and Ann Brown have developed methods of identifying learning-dis-

abled students and, on the basis of this assessment, providing instructional programs designed to ameliorate their learning problems (see Campione & Brown, 1987; Palinscar, Brown, & Campione, 1990).

The Campione-Brown approach utilizes a graded series of problem-specific hints that instruct the child in how to solve a task by focusing on the rules or principles involved in problem solution. The type and number of hints required serve as estimates of the child's learning potential in the domain assessed. At this level of analysis, the important measure is not the gain the individual makes as a result of the dynamic assessment, but how much aid is required for the child to reach a particular level of performance. Following the training (provision of hints), a posttest is administered with the pre- and postmeasure difference taken as an index of instructional gain.

An important element of Campione and Brown's approach is their provision for "transfer," that is, the ability of the learner to use the knowledge gained in the instructional setting to solve similar (near transfer) or different (far transfer) problems. The ability to transfer newly acquired skills to novel situations is important since it is a good predictor of how responsive the child will be to instruction (Farrara, Brown, & Campione, 1986).

Carlson and Wiedl's Testing-the-Limits Approach. Convinced that traditional methods of intelligence assessment systematically underestimate the mental capabilities of many children, Jerry Carlson and Karl Heinz Wiedl developed methods of assessment that involve modifications *within* the testing procedure. They assume that performance levels at or near the limits of the individuals' potentials provide the best indicators of their cognitive competence. In addition, they believe that such measures enhance the accuracy of measurement and improve the test's construct validity.

Carlson and Wiedl isolated a number of personal and/or noncognitive factors that negatively affect performance on complex cognitive tasks. These include anxiety, lack of motivation, impulsive responding, poor ability to plan, and lack of awareness of strategies useful in problem solution. The methods of assessment that were shown to ameliorate the negative effects of the performance-reducing factors involved active ver-

balization on the part of the individual taking the test and elaborated feedback provided by the examiner. The measures used include those with the highest *g* loadings, such as the RAVEN PROGRESSIVE MATRICES and the Cattell Culture-Fair Test.

In contrast to the traditional test-train-retest paradigm, the Carlson-Wiedl approach avoids statistical problems related to the measurement of change, and, in addition, is easy for the practitioner to use. In a series of investigations, Carlson and Wiedl have shown that their methods are effective with mildly retarded children (the lower threshold mental age appears to be about 6), deaf children, and culturally disadvantaged and ethnic minority children. In addition, their approach has important implications for teaching methodology where assessment and instruction can be theoretically and practically linked. (For a review, see Carlson & Wiedl, 1979, 1980, 1991, 1992.)

Guthke's Learning Test Approach. To improve the early diagnosis of children with potential learning difficulties, Jürgen Guthke has developed a number of learning tests (Guthke, 1977, 1992; Guthke & Wingenfeld, 1991), an example of which is the "Reasoning Learning Test" (Guthke, Jäger, & Schmidt, 1983). After initial administration, the children are provided with training that involves manuals designed to teach problem-solving strategies. The posttest, composed of items parallel to those on the pretest, is then administered, and the degree to which the children profit from the training is assessed.

Guthke has demonstrated that intelligence, especially fluid intelligence, is highly trainable. More important, the posttest scores are *not* predicted by pretest performance—evidence for the effectiveness of the training programs. In addition, Guthke has found that his learning tests correlate highly with creativity tests and componential aspects of intelligence not assessed by IQ measures. The latter include reaction time to both elementary and complex cognitive tasks as well as short-term-memory measures.

Feuerstein's Learning Potential Assessment Device. Shaped by his experience in assessing the mental abilities of unschooled North African immigrants in Israel and his belief that the poor performance of many individuals results from cultural and social deprivation, Reuven Feuerstein developed a sys-

tem of dynamic assessment called the LEARNING POTENTIAL ASSESSMENT DEVICE (LPAD). The model, described in detail in Feuerstein (1979), is based on the view that the goal of assessment should be to determine the potential of individuals to modify their cognitive abilities. For Feuerstein, the modification of cognitive abilities means not only facilitating performance within the person's zone of proximal development but also creating new structures that facilitate adaptation to new and complex situations.

The fundamental assumption behind the LPAD is that restructuring cognitive abilities through instruction provides the best index of an individual's cognitive potential and the degree to which that person can profit from future instruction. Through mediated learning experiences, that is, learning experiences in which learning sets are developed through direct intervention, inadequate or less than optimal performance can be improved and the degree of structural change assessed.

The usual LPAD comprises eight tests, some of which are commonly known (for example, a variant of the Raven Progressive Matrices). Test selection depends on the goals of the assessment and is made with a view to selecting an optimal level of complexity, that is, neither too difficult nor too easy. Since the purpose of the LPAD is to explore how the individual changes as a result of intervention, consistency of administration and reliability is sacrificed in favor of flexibility and of tailoring the testing/mediational learning situation to the needs of the individual.

RESEARCH ON DYNAMIC ASSESSMENT

A significant amount of research has been done on dynamic or interactive assessment. Lidz (1987) and Haywood and Tzuriel (1992) contain extensive reviews of this literature. On the basis of findings involving a variety of dynamic assessment models, the following conclusions may be drawn:

1. Intervention during assessment improves cognitive functioning independent of the effects of practice.
2. The most powerful intervention techniques are those which involve overt verbalization on the part

of the individual being tested and/or elaborated feedback provided by the examiner.

3. The greatest gains attributable to intervention are for lower-functioning individuals beyond a threshold mental age of approximately six years.
4. Predictive validities of tests administered using dynamic procedures are maximized if the method of assessment and method of instruction are matched.

BIBLIOGRAPHY

- BORTNER, M., & BIRCH, H. G. (1969). Cognitive capacity and cognitive competence. *American Journal of Mental Deficiency, 74*, 735-744.
- BRANSFORD, J. D., DELCLOS, V. R., VYE, N. J., BURNS, S., & HASSELBRING, T. S. (1987). Approaches to dynamic assessment: Issues, data and future directions. In C. S. Lidz (Ed.), *Dynamic assessment*. New York: Guilford Press.
- BUDOFF, M. (1967). Learning potential among institutionalized young adult retardates. *American Journal of Mental Deficiency, 72*, 404-411.
- BUDOFF, M. (1970). Learning potential: A supplementary procedure for assessing the ability to reason. *Acta Paedopsychiatrica, 37*, 293-309.
- BUDOFF, M. (1987). Measures for assessing learning potential. In C. S. Lidz (Ed.), *Dynamic assessment*. New York: Guilford Press.
- BUDOFF, M., & CORMAN, L. (1974). Demographic and psychometric factors related to improved performance on the Kohs learning potential procedure. *American Journal of Mental Deficiency, 78*, 578-585.
- BUDOFF, M., & FRIEDMAN, M. (1964). "Learning potential" as an assessment approach to the adolescent mentally retarded. *Journal of Consulting Psychology, 28*, 434-439.
- BUDOFF, M., & GOTTLIEB, J. (1976). Special class EMR children mainstreamed: A study of an aptitude (learning potential) \times treatment interaction. *American Journal of Mental Deficiency, 81*, 1-11.
- CAMPIONE, J. C., & BROWN, A. L. (1987). Linking dynamic assessment with school achievement. In C. S. Lidz (Ed.), *Dynamic assessment*. New York: Guilford Press.
- CARLSON, J. S., & WIEDL, K. H. (1979). Toward a differential testing approach: Testing-the-limits employing the Raven matrices. *Intelligence, 82*, 559-564.
- CARLSON, J. S., & WIEDL, K. H. (1980). Applications of a dynamic assessment approach in intelligence assessment:

- Empirical results and theoretical formulations. *Zeitschrift für Differentielle und Diagnostische Psychologie*, 1, 303–318.
- CARLSON, J. S., & WIEDL, K. H. (1991). The dynamic assessment of intelligence. In H. C. Haywood & D. Tzuriel (Eds.), *Interactive assessment*. New York: Springer.
- CARLSON, J. S., & WIEDL, K. H. (1992). Principles of dynamic assessment: The application of a specific model. In J. S. Carlson (Ed.), *Cognition and educational practice*. Greenwich, CT: JAI Press.
- EMBRETSON, S. (1987). Toward development of a psychometric approach. In C. Lidz (Ed.), *Dynamic assessment*. New York: Guilford Press.
- FARRARA, R. A., BROWN, A. L., & CAMPIONE, J. C. (1986). Children's learning and transfer of inductive reasoning rules: Studies in proximal development. *Child Development*, 57, 1087–1099.
- FEUERSTEIN, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device: Theory, instruments, and techniques*. Baltimore: University Park Press.
- FEUERSTEIN, R. (1980). *Instrumental enrichment: An intervention program for cognitive modifiability*. Baltimore: University Park Press.
- GUTHKE, J. (1977). *Zur Diagnostik der intellektuellen Lernfähigkeit*. Stuttgart: Klett.
- GUTHKE, J. (1992). Learning tests: The concept, main research findings, problems and trends. In J. S. Carlson (Ed.), *Cognition and educational practice*. Greenwich, CT: JAI Press.
- GUTHKE, J., JÄGER, C., & SCHMIDT, I. (1983). *Lerntestbatterie "Schlussfolgerndes Denken" (LTS)*. Berlin: Psychodiagnostisches Zentrum der Humboldt-Universität zu Berlin.
- GUTHKE, J., & WINGENFELD, S. (1991). The learning test concept: Origins, state of the art, and trends. In H. C. Haywood & D. Tzuriel (Eds.), *Interactive assessment*. New York: Springer.
- HAYWOOD, H. C., & TZURIEL, D. (Eds.). (1992). *Interactive assessment*. New York: Springer.
- LIDZ, C. (Ed.). (1987). *Dynamic assessment*. New York: Guilford Press.
- LIDZ, C. (1992). Dynamic assessment: Some thoughts on the model, the medium, and the message. In J. S. Carlson (Ed.), *Cognition and educational practice*. Greenwich, CT: JAI Press.
- MINICK, N. (1989). Implications of Vygotsky's theories for dynamic assessment. In C. Lidz (Ed.), *Dynamic assessment*. New York: Guilford Press.
- PALINSCAR, A. S., BROWN, A. L., & CAMPIONE, J. C. (1990). Dynamic assessment. In H. L. Swanson (Ed.), *Handbook of assessment of learning disabilities*. Greenwich, CT: JAI Press.
- SCARR, S. (1981). Testing for children: Assessment and the many determinants of intellectual competence. *American Psychologist*, 36, 1159–1166.
- VYGOTSKY, L. S. (1935/1978). Interaction between learning and development. In L. S. Vygotsky, *Mind in society: The development of higher psychological processes* (M. Cole, V. John-Steiner, S. Scribner, & E. Souberman, Eds. & Trans.). Cambridge: Harvard University Press.

JERRY S. CARLSON

DYSCALCULIA *Dyscalculia* means impairment of mathematical ability (in contrast to *dyslexia*, which means impairment of reading ability). Although research into mathematical disabilities has been limited, it has been estimated that more than 5 percent of the population experience mathematical deficits severe enough to be considered dysfunctional. Mathematical disorders fall into two general categories, based on the extent of dysfunction and the age at which the dysfunction was acquired. While there is disagreement among professionals regarding the specific distinctions between acalculia and dyscalculia, it is generally recognized that the term *dyscalculia* refers to mathematical deficiencies acquired during adulthood by an individual who formerly was proficient in arithmetic. More specifically, acalculia is considered to be present in an adult when mathematical functions have been affected without an overlapping reduction in general mental abilities. The term *dyscalculia* generally refers to a limited or specific impairment of mathematical ability or an impairment of a developmental nature. In developmental dyscalculia, hereditary or congenital factors are presumed to affect brain functioning, resulting in a general retardation of mathematical abilities. Performance is significantly below what would be expected of a child with a normal level of intelligence. Both adult and child disorders involve difficulty with counting; both can also be characterized by other deficits, such as inability to identify or execute mathematical functions (e.g., subtraction and addition) or inability to properly identify numbers and their symbology (value).

SUBTYPES AND SYMPTOMS

A variety of dyscalculic or acalculic subtypes have been described; however, a consensus has not yet emerged regarding the most appropriate diagnostic terminology or method of classification. It is therefore likely that considerable overlap exists among these types.

Acquired, or postlesional, acalculia is an acquired dysfunction in mathematical abilities that results from brain damage following a previously normal level of arithmetic capabilities.

Pure acalculia and anarithmetria (an inability to count or use numbers) are disorders of mathematical ability that occur in the absence of perceptual deficits. Thus this subtype is limited to isolated disorders of mathematical processing ability.

Alexia for numbers (loss of the ability to recognize numbers) is present when the individual cannot correctly recognize or write numbers. In most cases, alexia for words or an inability to recognize letters is also present. Alexia for numbers has also been referred to as aphasic acalculia.

Spatial acalculia occurs when the individual has difficulty perceiving or organizing arithmetic problems visually. For example, the individual may ignore the right half of a set of numbers when executing an operation either visually or in writing. However, the same calculations can usually be executed mentally following an oral presentation. Both alexia for numbers and spatial acalculia involve visual perceptual deficits.

Oligocalculia (gross or generalized dyscalculia) is a more general or diffuse deficiency affecting a wide range of mathematical skills. Oligocalculia is generally associated with mental retardation, whereas other forms of acalculia occur in the absence of concurrent mental deficiency.

Verbal dyscalculia refers to cases in which computations can be accurately performed, but the individual is unable to *name* digits, symbols, terms, amounts, or mathematical processes. One of the most commonly noted symptoms of verbal dyscalculia is the inability to perform simple counting tasks.

Practognostic (elaborate) dyscalculia specifies a deficit in the ability to apply abstract mathematical symbols to objects or drawings of objects. The individual is unable to quantify or order quantities of objects

from smallest to largest on the basis of shape, size, or spatial details. The literature occasionally refers to this as a form of apraxic dyscalculia.

An inability to conceptualize mathematical functions that involve symbols—such as operational signs (e.g., a plus or minus sign) or components of equations—is the predominant distinguishing symptom of lexical dyscalculia. This form of dyscalculia may be paired with other subtypes, or may be seen in conjunction with other forms of reading disabilities, such as dyslexia (a level of reading ability markedly below that expected on the basis of the individual's level of overall intelligence or skills). Lexical dyscalculia is occasionally referred to as numerical dyslexia. It is hypothesized that lexical dyscalculia is caused by a deficit in visual-spatial processing.

An individual manifesting graphical dyscalculia or graphical dyslexia is unable to convert an orally presented number into its written representation, to rewrite words in numerical or digit form, or to copy written figures. The individual can convert a written number into word form, however.

Ideognostic dyscalculia/dysymbolia is an inability to understand basic mathematical concepts and their relation to other concepts. More specifically, individuals with ideognostic dyscalculia are unable to complete even simple addition mentally.

Operational dyscalculia is an inability to manipulate mathematical processes accurately. The resulting mathematical solutions either will be incorrect because of the use of inappropriate mathematical rules or, if accurate, will have been derived incorrectly. Operational dyscalculia is often considered the most difficult of the subtypes to diagnose because the logic used to complete an operation may be difficult to follow.

Pseudodyscalculia is a deficit in mathematical functioning due to limited education or a poor learning history. When an individual lacks motivation to learn, has had poor schooling, or has an illness that impedes learning of mathematical skills, the resulting syndrome may appear quite similar to true dyscalculia/acalculia, although it is not associated with any acquired or developmental cerebral dysfunction.

A variety of other deficiencies, which are not necessarily linked to specific subtypes of dyscalculia/acalculia, can produce computational difficulties. Paraphasic substitutions are those errors in which one

number is inappropriately substituted for another in mathematical operations. Perseveration errors occur when the individual persistently repeats the use of a number or a particular arithmetic operation when a different response is required. In digit reversals (a common error), numbers are written upside down (e.g., 2 is written as 5) or the order of numerals is reversed (e.g., 327 is written as 723). In spatial organization errors, an individual has difficulty with correctly arranging numbers, as in addition of columns of numbers. Spatial organization errors also occur if an individual executes an operation correctly but begins at the wrong place (e.g., if he or she begins the addition or subtraction of a column of numbers from the left rather than the right). Visual detail errors occur when an individual reads a mathematical symbol incorrectly, executing a division instead of a subtraction operation, or multiplication rather than addition. Procedural errors result from incorrectly performing standard mathematical operations (e.g., neglecting a necessary step or scrambling the proper sequence of operations). Graphomotor problems are essentially due to an inability to write numerals properly, resulting in illegible or reversed numerals. Judgment and reasoning errors occur when an individual is unable to identify the erroneous result of an operation as obviously incorrect.

BRAIN LOCALIZATION

Both acalculia and dyscalculia result from localized lesions of the brain. Most commonly, these lesions occur in the left cerebral hemisphere. More specifically, most cases of acalculia have been noted as resulting from left parietal lobe lesions. Lesions in these areas can also result in deficiencies in reading and spelling. Lesions in the right hemisphere may also result in mathematical deficiencies, primarily of the spatial type. Bilateral lesions have also been associated with some forms of acalculia.

Acquired acalculia results from focal lesions sustained after basic mathematical skills have been developed. Developmental dyscalculia results from genetic or congenital deficiencies, which are presumed to involve the brain centers noted for mathematical processing. Developmental dyscalculia impedes the learning process.

EVALUATION AND DIAGNOSIS

The child's age and level of general intellectual functioning must be taken into account before a diagnosis of dyscalculia can be made. Dyscalculia can be diagnosed only when mathematical abilities are significantly reduced in comparison to the individual's overall level of mental functioning. Since the development of abstract expression is required for mathematical learning, and since some exposure to mathematics is required, this evaluation process is generally not attempted until the later primary school years. Examination involves the assessment of intelligence and mathematics achievement by means of standardized, individually administered psychological tests. The Wechsler and Stanford-Binet intelligence tests are widely used for this purpose. The Wide Range Achievement, Woodcock-Johnson, and Key Math tests are commonly administered measures of mathematical skills.

A variety of neuropsychological techniques are effective in discriminating between normally intelligent children with mathematical deficiencies and normally intelligent children with adequate mathematical skills. These tools are also used to determine if mathematical deficiencies are the result of poor learning history or are neurologically based.

The Rey-Osterrieth Complex Figure test is essentially a visual-motor test; it requires the subject to accurately copy a line drawing of a complex geometric figure. Thus, accurate perception is a key factor in the subject's ability adequately to carry out the task. Evaluation of the reproduction is based on the points earned for each portion of the model reproduced correctly and the amount of time it takes the subject to initially start drawing. It has been suggested that this test is particularly useful as an aid in the diagnosis of mathematical problems related specifically to spatial deficits. Kalkulia III is a test that requires the subject to differentiate patterns within a larger pattern of variously colored circles. It involves the ability to quantify shapes, convert quantities into numbers, and perform addition. The Numerical Triangle test consists of a two-dimensional array of single-digit numbers. Entries for one dimension of the array and simple addition rules for completing the other dimensions of the array are provided. This test assesses the subject's ability to

perform addition and to follow directions on mathematical tasks. Children who demonstrate a subtype of dyscalculia that involves deficits in spatial ability, such as graphic dyscalculia, will have noticeable difficulties on this assessment technique.

Digit span is another potentially useful technique. The examiner presents from three to seven digits to the subject at a rate of one per second. The subject is required to repeat the sequence of numbers back to the examiner in the correct order. A technique called serial sevens requires subjects successively to subtract 7, starting with 7 from 100. It is useful as a crude indicator of deficiencies in counting abilities.

Although acalculia can occur in isolation, it can often be a symptom of other neurological disorders. For example, acalculia may be a symptom of Gerstmann's syndrome, Alzheimer's disease, or cerebrovascular accidents. Acalculia and dyscalculia are often mistaken for dyslexia because some of the manifestations, such as reversal of digits, are similar to dyslexia. These disorders may be linked, since hemispheric lesions of the left parietal lobe that result in acalculia or dyscalculia often cause dyslexia as well. Although individuals with low intelligence scores may appear to be acalculic, the diagnosis technically excludes mathematical deficits that coincide with general mental disabilities or impaired intelligence.

PROGNOSIS AND TREATMENT

Only limited information is available at present concerning the outcome of mathematical-processing disorders. Prognosis is in part dependent upon the etiology of the disorder. The first step in remediating the math skills of dyscalculic children is to determine the limits of their mathematical understanding. Any difficulties in the integration of sensory and motor functions should be identified. Before remediation can be attempted, it is necessary to determine what faulty mathematical problem-solving strategies and logic are currently being used by the child. The degree to which the child is able to listen, communicate effectively, and comprehend must also be evaluated. Finally, an analysis of the child's reading, writing, and math skills is necessary.

Remediation for dyscalculia focuses on identifying and strengthening auxiliary and complementary skills.

Many dyscalculic deficiencies involve difficulties in symbolic communication or the ability to associate linguistic and symbolic representations of a mathematical concept. In other words, many dyscalculic children are unable to perform the processes that intervene between the verbal explanation of a mathematical concept and the application or execution of these concepts with abstract numerical symbols. Remedial efforts are frequently directed at generalizing existing skills and strengths to mathematical tasks for this reason. Since these strengths vary from child to child, the remediation process is highly individualized. A consistently structured, highly verbal learning environment is necessary for dyscalculic children to develop alternative skills. Examples of techniques that are used in this intervention process include rehearsal, drills, training with appropriate tools (e.g., calculators), and instruction in procedural sequences. Behavioral and environmental interventions are frequently useful adjuncts to the educational remediation of specific arithmetical deficits. These interventions can include development of verbal skills and expressiveness, expanded learning activities in areas outside of mathematics, and parental and peer training. Since dyscalculic children demonstrate substantial organizational deficits, they may exhibit confusion or lack of motivation—especially while learning or practicing general problem solving. This is not necessarily the result of inattention, but rather a secondary effect of the dyscalculic deficiencies. The goal of the treatment process is not to eliminate the deficiency, but to allow the individual to function at the highest level possible with his or her available skills.

(See also: MENTAL DISABILITIES.)

BIBLIOGRAPHY

- ARDILA, A., & ROSSELLI, M. (1990). Acalculias. *Behavioral Neurology*, 3, 39–48.
- BOLLER, F., & GRAFMAN, J. (1983). Acalculia: Historical development and current significance. *Brain and Cognition*, 2, 205–223.
- CORBETT, A. J., MCCUSKER, E. A., & DAVIDSON, O. R. (1986). Acalculia following a dominant-hemisphere subcortical infarct. *Archives of Neurology*, 43(9), 964–966.

- FERRO, J. M. (1990). Anatomical correlates of acalculia following left-hemisphere stroke. *Behavioral Neurology*, 3, 7–17.
- FILSKOV, S. B., & BOLL, T. J. (1981). *Handbook of clinical neuropsychology*. New York: Wiley Interscience.
- FLETCHER, J. M., & LOVELAND, K. A. (1986). Neuropsychology of arithmetic disabilities in children. *Focus on Learning Problems in Mathematics*, 8(2), 23–40.
- FLINTER, P. F. (1979). Educational implications of dyscalculia. *Arithmetic Teacher*, 27(7), 42–46.
- GORDON, N. (1992). Children with developmental dyscalculia. *Developmental Medicine and Child Neurology*, 34, 459–463.
- HECAEN, H., & ALBERT, M. (1978). *Human neuropsychology*. New York: Wiley Interscience.
- HEILMAN, K. M., & VALENSTEIN, E. (Eds.). (1979). *Clinical neuropsychology*. New York: Oxford University Press.
- JOSEPH, R. (1990). *Neuropsychology, neuropsychiatry, and behavioral neurology*. New York: Plenum.
- KOLB, B., & WHISHAW, I. (1990). *Fundamentals of human neuropsychology*. New York: Freeman.
- KOSC, L. (1974). Developmental dyscalculia. *Journal of Learning Disabilities*, 7(3), 164–177.
- KOSC, L. (1981). Neuropsychological implications of diagnosis and treatment of mathematical learning disabilities. *Topics in Learning and Learning Disabilities*, 1(3), 19–30.
- KOSC, L. (1986a). Basic forms of developmental dyscalculia. *Focus on Learning Problems in Mathematics*, 8(3–4), 55–61.
- KOSC, L. (1986b). Developmental dyscalculia. *Focus on Learning Problems in Mathematics*, 8(3–4), 48–53.
- KOSC, L. (1986c). Diagnosing mathematical dysfunctions in children. *Focus on Learning Problems in Mathematics*, 8(3–4), 67–70.
- KOSC, L. (1986d). Diagnostic process and tests of dyscalculia: Kaculia III, Rey-Osterieth complex figure test, number triangle. *Focus on Learning Problems in Mathematics*, 8(3–4), 71–83.
- KOSC, L. (1986e). Terminology in the field of developmental dyscalculia. *Focus on Learning Problems in Mathematics*, 8(3–4), 63–65.
- KOSC, L. (1986f). The correction of dysfunction in children's mathematical abilities. *Focus on Learning Problems in Mathematics*, 8(3–4), 93–100.
- SHALEV, R. S., WEIRTMAN, R., & AMIR, N. (1988). Developmental dyscalculia. *Cortex*, 24, (555–561).
- SHALLICE, T. (1988). *From neuropsychology to mental structure*. New York: Cambridge University Press.
- SHARMA, M. C. (1986). Dyscalculia and other learning problems in arithmetic: A historical perspective. *Focus on Learning Problems in Mathematics*, 8(3–4), 7–45.
- SVIEN, K., & SHERLOCK, D. (1979). Dyscalculia and dyslexia. *Bulletin of the Orton Society*, 29, 269–276.
- TEMPLE, C. M. (1989). Digit dyslexia: A category-specific disorder in developmental dyscalculia. *Cognitive Neuropsychology*, 6(1), 93–116.
- TEMPLE, C. M. (1991). Procedural dyscalculia and number fact dyscalculia: Double dissociation in developmental dyscalculia. *Cognitive Neuropsychology*, 8(2), 155–176.
- WALSH, K. W. (1978). *Neuropsychology: A clinical approach*. New York: Churchill Livingstone.

WILEY MITTENBERG
MICHELLE L. BENGTSON

DYSLEXIA Dyslexia is a language disorder characterized by difficulty in decoding the written word into its spoken form. Usually it first becomes apparent in the early school years, when youngsters are taught to read and spell. Both of these academic skills require children to learn the code associating the sounds of words with their written symbols. Thus, dyslexia is often thought of as a specific learning disability in the areas of reading and spelling. Although reading and spelling are the primary areas of academic difficulty, some children with dyslexia also have difficulties with mathematics.

The term is derived from the Latin, *dys* meaning “difficult” and *legre* meaning “to read,” and from the Greek, *lexis* meaning “speech.” Generally, it is considered a developmental disorder caused by genetic, gestational, or prenatal factors. Dyslexia also may be an acquired disorder that occurs after a normal period of development but is due to injuries to a child's immature brain. The brain-injury-related loss of reading skills in an adult who could read prior to the injury is called alexia. Knowledge gained from studies of individuals with alexia has contributed significantly to what currently is known about dyslexia.

Dyslexia is a lifelong condition, with consequences that often affect the individual's educational progress, self-esteem, employment, and relationships with other people. Currently, no treatment exists to eliminate dyslexia. However, early and sustained management

TABLE 1
Characteristics of dyslexia

Preschool and kindergarten	
Primary	No primary characteristics because dyslexia is not formally identified until one or two years of education have been completed
Associated	Mild speech delay; talking less than other children do; word-finding problems; misarticulations (e.g., <i>thire</i> for <i>fire</i>), syllable missequencing (e.g., <i>gunjel</i> for <i>jungle</i>); problems with remembering verbal sequences (such as "Get your coat, put on your socks and shoes, and turn off the TV"); difficulty in learning nursery rhymes; difficulty in learning the names of letters, colors, or numbers
Secondary	Generally not observed at this age because the child is unaware or only mildly aware of any difficulties
School age	
Primary	Difficulty in learning letter-sound correspondences; difficulty in sounding out words; slow, halting, and often inaccurate oral reading; difficulty in learning to read; difficulty in learning to spell; slow and halting writing
Associated	Letter and number reversals, particularly among children 9 years or older (e.g., <i>bog</i> for <i>dog</i>); incorrect use of function words (e.g., <i>on</i> for <i>over</i>); difficulty in memorizing basic math facts; difficulty in learning the names of the symbols of arithmetic operations (such as plus and minus signs)
Secondary	Symptoms of stress and failure, such as anxiety, depression, low self-esteem, tearfulness, irritability, hostility, refusal to attend school, headaches, or stomach aches
Adolescent and adult	
Primary	If the individual has been unable to compensate for dyslexia, primary characteristics like those at school age; if the individual has compensated, less severe characteristics (some symptoms may even appear absent to the casual observer); slow reading speed; poor comprehension on timed reading tests; misreading of function words (e.g., <i>the</i> for <i>a</i>); visual reading errors (e.g., reading <i>dungeon</i> for <i>dragon</i>); misspellings that are phonetically accurate (e.g., <i>nacher</i> for <i>nature</i>)
Associated	Difficulty in completing homework; incorrect use of function words (e.g., <i>the</i> for <i>a</i>) and suffixes (e.g., <i>presenting</i> for <i>presently</i>); difficulty in remembering complex verbal sequences
Secondary	Underemployment relative to level of ability; shame; embarrassment; depression; anxiety; dissatisfaction

and support can minimize its potentially negative consequences.

The fields of education, psychology, and medicine have contributed to what is currently understood about dyslexia, and members of these disciplines have slightly different points of view as to its existence as a specific syndrome, its definition, its causes, and its management. For example, some professionals deny that dyslexia exists, arguing that the problem reflects an external deficit in the method of instruction rather than a condition that is internal to the individual. Other professionals object to the term *dyslexia* because it labels people, which presumably causes negative effects, such as teasing or lowered expectations for success. Some professionals use the term loosely to refer to all kinds of learning disabilities; still others use the term to refer only to acquired or developmental disorders in reading and spelling. These different points of view result in a good deal of controversy about the specific definition of dyslexia and how it is identified in individual cases.

Typically, dyslexia is defined and identified with exclusionary rather than inclusionary rules. These definitions tell us what dyslexia is not, rather than what dyslexia is. For example, the most widely accepted definitions exclude people with below-average intelligence, mental retardation, vision or hearing impairments, and/or cultural and educational deprivation. The only characteristic that is diagnostic of dyslexia in these definitions is the individual's significant difficulty in learning to read and spell, relative to expectations based on age and intelligence.

Exclusionary definitions have many problems; perhaps the biggest is the fact that children must fail at reading and spelling before they can be identified as dyslexic. If these early failures are not identified and managed, they may establish an increasingly downward spiral of diminished self-esteem, with reduced motivation to learn, followed by further failure. Consequently the child may develop a set of negative emotions toward learning and school, which further compounds their difficulties. Early detection and management of dyslexia are needed to prevent this downward spiral. Fortunately, recent research has focused on identifying the characteristics of preschool-age children who later developed dyslexia. In the future, these characteristics will be useful in developing an accepted

definition based on inclusionary rules. Until then exclusionary definitions of dyslexia will remain.

WHAT ARE THE CHARACTERISTICS OF DYSLEXIA?

Because learning to read is a developmental process, we expect different levels of reading skills at different ages, and the characteristics of dyslexia differ somewhat across ages. For example, although it may be sufficient for 5-year-old children to know only the associations between letters of the alphabet and their sounds, we expect older children to blend these sound into words, and we expect still older children to comprehend unfamiliar words and text.

As stated earlier, many specialists believe that dyslexia cannot be identified until the individual is old enough to have had some educational experience, usually by the end of the first or second grade. Nevertheless, certain early characteristics seen in preschool-age children are associated with subsequent dyslexia. The characteristics of dyslexia are listed in Table 1. They are grouped by two factors: age and symptom level.

Consistent with the exclusionary definition of dyslexia, the primary characteristics are difficulty in learning to read and spell. Associated symptoms reflect the difficulty with language processes that presumably underlies dyslexia. Secondary symptoms reflect individuals' emotional and physical reactions to their environment and to unrealistic or negative expectations on the part of their teachers and others. Secondary symptoms often result in a referral to a specialist for evaluation.

HOW IS THE INDIVIDUAL WITH DYSLEXIA IDENTIFIED?

Identifying whether someone has dyslexia is a complex process. Specialists often interpret the exclusionary definition differently. Consequently, they do not agree on just one standard set of evaluative procedures. For example, consider the statement that the individual with dyslexia has "significant difficulty learning to read and spell." Some specialists interpret this to mean that individuals must score two or more years behind their grade placement on standardized reading and spelling tests (e.g., a third-grader scoring

at the first-grade level). Others interpret this to mean that individuals must score 15 points or more below their overall intellectual ability on these tests. These two different interpretations of the same phrase may result in different professional judgments as to whether someone has dyslexia.

Specialists try to minimize differences in professional judgment by gathering many pieces of information and test scores from different sources. Then they evaluate this information for consistency. Dyslexia is identified if several sources, such as test results and family history, are consistent with the exclusionary criteria and characteristics of the condition. Sometimes evaluations for dyslexia are a team effort; in other instances a single specialist, usually a psychologist, conducts the evaluation.

Generally, the public education system uses a team of specialists to evaluate students. The evaluation may include reviews of academic records, interviews, taking a family history, a physical examination, and individual testing. Typically, the team includes the parents, the teacher, and a psychologist. Some teams also include a speech and language pathologist, special educator, occupational therapist, physical therapist, physician, nurse, and/or social worker. The psychological assessment is critical because it provides several pieces of information, including an evaluation of the student's overall intellectual ability. The psychological assessment may also include information about the child's reading and spelling levels, and the mental processes that presumably underlie these skills, such as memory. Evaluations from other specialists are necessary to determine whether factors such as sensory impairments (e.g., poor hearing) or insufficient educational opportunity may be the cause of the reading and spelling difficulty. If these other factors are eliminated and the child has at least average intellectual ability but gets scores below this ability on reading and spelling tests, the child is considered to have dyslexia.

Currently most educational and psychological practices prefer classifications based on 15- or 16-point differences between ability and achievement test standard scores. Classifications based on two-year differences between grade placement and earned grade-level scores are less accurate and tend to exclude high-ability students with dyslexia. Among children 7 to 8 years old and younger, the diagnosis of dyslexia may be

based on the presence of primary and secondary symptoms, even if test-score differences are not sufficiently large. Similarly, test-score differences may not be large enough among older adolescents and adults who have learned strategies to cope with their difficulties. Consequently, in this age group, dyslexia may be diagnosed on the basis of past school performance, family history, and current characteristics.

HOW COMMON IS DYSLEXIA?

Most authorities would agree that dyslexia is the most common and best understood learning disability. Estimates of exactly how common it is vary, depending on how researchers define dyslexia and how they identify whether individuals have dyslexia. Estimates range from as low as 5–10 percent to as high as 20–30 percent. However, the higher estimates include children with reading disabilities stemming from below-average intelligence or limited cultural and educational opportunities. Consequently, the more conservative estimate is 10 percent. Researchers typically observe that dyslexia is more common in boys than in girls, citing a ratio of three or four boys to every girl. The properties of tests and procedures used to identify individuals with dyslexia may account for some differences between the rates of dyslexia in boys and girls. Girls also may learn to cope with dyslexia sooner and more successfully than boys, presumably because girls tend to have better language skills than do boys. Social explanations for these differences suggest that boys with dyslexia are more likely to come to the attention of teachers and specialists because they are more difficult to manage than girls. Still other explanations propose genetic differences or sex-linked hormonal differences in the child's prenatal, intrauterine environment.

Recent, scientifically sound research showing that dyslexia runs in families favors biological explanations of gender differences. In the Colorado Family Reading Study, boys born to a parent with dyslexia ran a 35–40 percent risk of developing dyslexia themselves, with a slightly higher risk if the dyslexic parent was the boy's father (40 percent) than if the dyslexic parent was his mother (35 percent). Girls born to a parent with dyslexia ran a 17–18 percent risk of developing dyslexia, regardless of the affected parent's gender. Boys born to parents who did not have dyslexia ran

only a 5–10 percent risk of developing the condition themselves, while girls born in similar circumstances ran a 1–2 percent risk.

WHAT CAUSES DYSLEXIA?

The cause of dyslexia is still unknown although many theories have been proposed and researched. It appears that no single cause exists, and that dyslexia may represent the final common end point of several different causal pathways. Explanations for dyslexia have been offered at the biomedical (i.e., brain structure, hormonal influence, genetics), neuropsychological (i.e., functional processes of the brain), and environmental levels. Ultimately, causal explanations must include the fact that learning to read is a developmental process that evolves as the child's brain matures and as the child practices the skill. The developmental model of dyslexia, however, is dependent on establishing a developmental model of normal reading. Because the most promising progress to date has been made in biomedical research, this article emphasizes biomedical explanations of dyslexia.

Biomedical Explanations. Several studies have confirmed that the metabolic activity and brain

structure of individuals with dyslexia are different from those of individuals who are able to read. Scientists recently used positron-emission tomographic (PET) scans to compare the brain's metabolic activity of adults with dyslexia since childhood to that of adults with lifelong normal reading skills. Study participants listened to an audiotape of cocktail-partylike noises and speech sounds. They were asked to respond to specific speech sounds that were embedded randomly on the tape. Adults with dyslexia made significantly more errors on this task and showed greater metabolic activity in areas of the right and left temporal lobes than did their normal-reading counterparts. Similar studies of children with dyslexia await further refinement of PET-scan measurement technology.

Although differences in brain metabolic activity found in the PET-scan study likely were due to differences in how the participants' brains developed, differences between individuals with dyslexia and normal readers also may be due to acquired factors. For example, brain autopsies of adults who acquired reading and spelling disorders after a normal period of development found damage in parts of the left temporal lobe and adjoining areas (see Figure 1). The left tem-

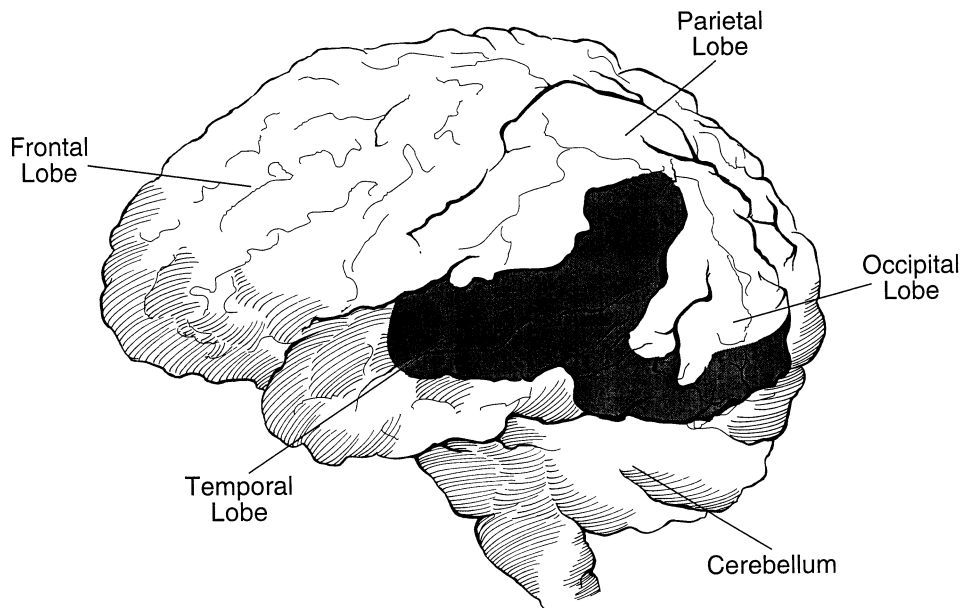


Figure 1

Areas of the left cortical hemisphere involved in basic reading processes

poral lobe is important for processing auditory and language-based information. It houses memory for language and associations between visual, auditory, motor, and sensory functions.

Anatomical studies have found that the left temporal area is larger than the same area on the right cortical hemisphere in most individuals. However, this type of asymmetrical brain structure is not found among many individuals with dyslexia. Microscopic studies of brain tissue from left temporal areas of individuals with developmental dyslexia have shown malformed and misarranged nerve cells. These tiny malformations and misarrangements cannot be detected by contemporary medical tests, such as computed tomography (CT) or magnetic resonance imaging (MRI) scans. Presumably they are due to disordered migration of brain cells early in the development of the fetus.

The Geschwind and Behan theoretical model attempts to explain some of these brain differences and proposes that dyslexia is due to an unborn child's increased exposure to the male hormone, testosterone, during gestation. The theory also tries to account for gender differences in the incidence of dyslexia, the disputed historical belief that dyslexia and left-hand preference are associated with each other, and the observed association between dyslexia and immune disorders in certain individuals. Presumably, increased testosterone slows the development of the immune system and makes the individual vulnerable to such disorders as asthma, eczema, and hay fever. Increased testosterone also may alter the course of fetal brain and brain structure development, particularly on the left side (left cortical hemisphere). In most individuals, right-hand preference and language-based activities necessary for reading are a function of the left cortical hemisphere. Altered development of the left hemisphere presumably causes both dyslexia and a shift in hand preference from right to left. Results from contemporary studies that investigated the associations among dyslexia, left-handedness, and immune disorders have been mixed, however. Consequently, researchers speculate that these conditions may occur together in only a small subgroup of families and individuals with dyslexia.

Apparently, the historical belief of an association between dyslexia and left-handedness may stem from

the association between these conditions among individuals who suffered brain injuries to the left hemisphere. Such injuries can cause both language dysfunctions, such as dyslexia, and a shift in hand preference from right to left. Thus, in this group, language, reading difficulties, and handedness are related because of the injury. Most contemporary researchers agree that except in this group, there is generally no association between left-handedness and dyslexia.

As suggested earlier by data in the Colorado Family Reading Study, there appears to be a genetic component to dyslexia. For example, among families at risk for dyslexia, a history of the disability predicted children's reading skills better than did environmental factors, such as socioeconomic status or exposure to reading and television. It is likely that several genes, rather than a single gene, contribute to dyslexia. Contemporary studies suggest that these genes affect specific language skills necessary for reading, rather than affecting reading itself. Currently, investigators are attempting to locate these genes on chromosomes 15 and 6. Some studies have confirmed a genetic locus on chromosome 15 in certain families. Studies that confirm genetic locus on chromosome 6 may account for the association between dyslexia and immune disorders, because certain areas on chromosome 6 possess genetic material responsible for various immunological functions.

Cognitive and Neuropsychological Explanations. Neuropsychological explanations have focused on pinpointing mental processing (i.e., cognitive) factors that presumably underlie dyslexia. Deficits in intellectual ability, visual processing, sequencing information, verbal memory, and language are some of the many cognitive factors thought to underlie dyslexia. Currently, we know that intellectual deficiency does not cause dyslexia. By definition, people with dyslexia are of at least average intelligence but read at a level below that expected for their ability.

Early research suggested that poor visual processing caused dyslexia. This notion is still commonly held by the general public, as exemplified by the belief that reversals in writing (e.g., *bog* for *dog*) and seeing words backward and upside down are the main features of dyslexia. In fact, reversals are common among all young children (through ages 7 and 8) and are not in and of themselves an indication of dyslexia.

Difficulties in sequencing information also may underlie dyslexia. For example, children might read *phelepant* for *elephant*. Reading requires the child to string together sequences of letters and their associated sounds in order to create words. Later children must sequence words to make sentences and sequence sentences to make paragraphs. Although sequencing deficiencies, particularly sequencing and organizing speech, are frequently found among dyslexics, these deficiencies do not appear to be the fundamental cognitive deficits that underlie dyslexia.

Verbal memory deficits are also common among people with dyslexia. Verbal memory is complex, however, and includes many processes. For example, some researchers argue that dyslexic children may lose track of word order, forgetting words they have previously decoded while struggling to decode the next, new word. Other researchers argue that children forget the sounds they must blend together to make words. These problems may reflect difficulties getting the information into memory (encoding) or difficulties getting it out of memory (retrieval). Some researchers argue that individuals with dyslexia have difficulty retrieving the spoken name of words, although they know the words because they have used them before. Although it is clear that people with dyslexia do have certain problems with verbal memory, current research suggests that verbal memory deficits are not the primary deficit in dyslexia.

Although some children with dyslexia do have difficulty with visual processing and memory, contemporary research suggests that in the majority of cases, the primary deficit actually involves certain language-processing skills. Specifically, most people with dyslexia have difficulty discriminating among the individual sounds of speech (phonemes). This deficiency makes learning the rules that associate the component parts of written words (graphemes) with phonemes very difficult. Consequently, people with dyslexia have tremendous difficulty sounding out words.

Guessing the target word is an attempt to cope with this difficulty. Guesses may be based on visual similarity (e.g., reading *horse* for *house*) or on the word's beginning grapheme-phoneme association and the content of the prior text (e.g., reading *ghost* for *ghoul*).

Researchers have attempted to define subgroups of dyslexia based on these types of reading errors, similar types of spelling errors, and other neuropsychological information. They have yet to agree whether there is just one or several subtypes of dyslexia. Promising research using computer models of learning, reading, and their disorders provides reasonably convincing evidence, however, that dyslexia (and alexia) likely have more than one single cause, and thus likely have many different subtypes.

Environmental Explanations. Environmental explanations for dyslexia are less well researched than biomedical or neuropsychological ones. Obviously, external insults, such as damage to the child's immature brain from a car accident or exposure to toxic substances (e.g., lead), can adversely affect later ability to learn. Additionally, some scientists have proposed that environmental insults to the fetus, such as exposure to pollutants, chemical toxins, or infectious agents, may be responsible for dyslexia.

Cultural factors also might account for dyslexia. For example, some researchers have proposed that dyslexia is a function of the type of writing system employed in different countries. They hypothesized that dyslexia is more common in countries with languages that use phonetically based written codes, such as English, than in countries with languages that use symbols that represent ideas without coding pronunciation (i.e., ideographs), such as Japanese kanji. Furthermore, some researchers suggested that dyslexic American students might learn to read if they were instructed in language and writing systems that used ideographs. Figure 2 shows two *kanji* (Japanese ideographs). Studies specifically designed to examine the rate of dyslexia across cultures have not been published yet. However, preliminary reports that examined school achievement among American, Japanese, and Taiwanese children found that each country had roughly the same number of children whose reading skills were behind their grade placement. This result does not support the notion that dyslexia is more common among languages with a phonetically based written code.

Apparently, oversimplifications of the differences between languages using ideographic versus phonic codes contributed to this notion. Most experts who fully understand the complex differences between

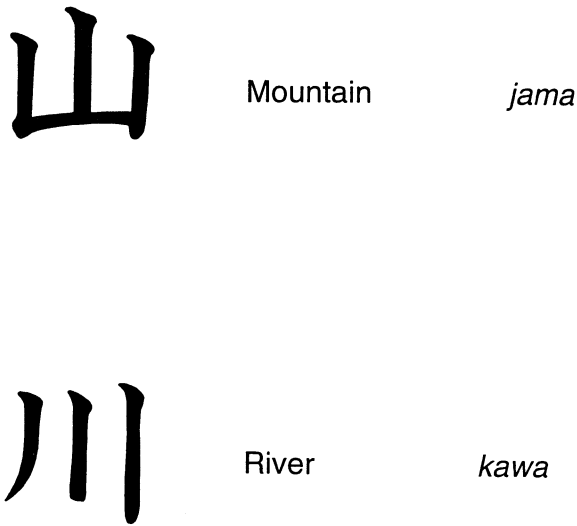


Figure 2
Two Japanese ideographs. Like all ideographs, Japanese characters (called *kanji*) code specific ideas; the top one means “mountain,” and the bottom one means “river.” Unlike written English words, Japanese *kanji* do not code pronunciation. Thus, despite the visual similarity of these two, they are pronounced differently: *jama* and *kawa*, respectively.

phonetically and ideographically based writing systems do not believe that American children with dyslexia can learn to read if instructed in a system that uses ideographs such as Chinese or Japanese.

Other possible environmental explanations for dyslexia include style of classroom instruction and family size. Different styles of instruction appear to promote different ways of thinking and learning. For example, educational researchers have shown that quality of learning improves if teachers use an active rather than passive style. Teachers with active styles expect and promote students’ verbal responses and dialogue. Teachers with passive styles tend to have fewer dialogues and give more lectures.

Large family size may affect cognitive development, also. Children in large families, where individual time with parents is limited, may have minimal exposure to reading and language games during their preschool years. These and other experiential factors presumably affect childrens’ phoneme awareness, cognitive skills, and later reading ability.

HOW CAN INDIVIDUALS WITH DYSLEXIA BE HELPED?

There are many contemporary interventions designed to improve the academic skills of individuals with learning disabilities such as dyslexia. For example, some require children to change their diets (e.g., Feingold diet, or eliminate sugar and caffeine), take vitamin and mineral supplements, be retaught how to crawl and walk, be trained in particular patterns of eye movements, or participate in play therapy. Few have been evaluated scientifically for their specific effectiveness with dyslexia. Although the medical field often is criticized for the length of time between the discovery of an effective treatment and its availability to the general public, the reverse is true in the field of education. Many interventions (treatments) are included in educational practice before their effectiveness has been evaluated carefully. Consequently, one educational psychologist warned her readership to, “read the fine print before accepting a ‘discovery’ as an academic or behavioral intervention” (Rooney, 1991, p. 134). Individuals who seek treatment for dyslexia would do well to heed this advice.

Many contemporary interventions have been based on unproven theories of the causes of dyslexia. Fortunately, scientific advances will result in improved interventions. One such advance is converging evidence that the awareness of speech sounds (phonological awareness) and the knowledge of associations between the written and spoken parts of speech (grapheme-phoneme associations) is critical to basic reading. Individuals with dyslexia are deficient in these areas and do not outgrow the deficit (although they can learn phonological awareness, albeit very slowly). As a result of these advances, many experts now agree that interventions for dyslexia must target the individual’s deficient phonological awareness and word-decoding skills (Pennington, 1991, p. 75). In other words, interventions should systematically teach children the associations between letters or letter patterns and their corresponding sounds. Accordingly, phonics-based approaches to reading are the preferred intervention for children with dyslexia. The success of these interventions appears to hinge on the child’s basic phoneme awareness (e.g., the ability to blend phonemes into

words or the ability to say the part of a word that remains after the initial phoneme is deleted). Consequently, phoneme awareness programs may be needed prior to starting a phonics-based reading program.

Formal Educational Programs. The field of education recognizes three major phonics approaches for use with dyslexic students. These are the Direct Instruction Reading program (DSTAR), the Orton-Gillingham approach, and the Spalding method (*The Writing Road to Reading*, Spalding & Spalding, 1969). The characteristics of these approaches are described in Table 2. Aaron and Joshi (1992, p. 130) critically reviewed these and other reading approaches. They concluded that of the three, the Spalding method had more critical and scientific evaluations of effectiveness with dyslexic students and that the results were good. Unfortunately, scientifically sound comparisons among different phonics-based reading programs with dyslexic students are not yet available, thus experts cannot state confidently whether one program is superior to another.

Older individuals with dyslexia, who have not had phoneme awareness training and phonics-based reading instruction, can improve their reading skills with these interventions. Yet they also need assistance with reading comprehension and learning effective study

habits, such as test-taking and listening-comprehension skills. Instructional approaches for these skills have been developed, but they have had little scientific evaluation of their effectiveness with dyslexic people.

Environmental Coping Techniques. People with dyslexia should not be penalized for their disability in home, school, or work environments. They have normal intellectual abilities and can learn. Teachers, employers, and parents must help children with dyslexia find effective ways to take in and later get out information. In this vein, changes to learning or work environments are helpful.

Attitudes individuals have toward themselves, as well as attitudes of others (e.g., teachers, parents, peers), are major environmental factors affecting individuals with dyslexia. Secondary emotional problems, such as diminished self-esteem and a feeling of failure, can impede an individual's progress. People with dyslexia are not lazy or stupid, although they commonly sense this perception from others. Adults have the responsibility to ensure that children respond to dyslexic classmates with kindness and compassion. Families must help identify and support recreational, leisure, and social activities that provide satisfaction for individuals with dyslexia. These activities can help build skills for the future. Additionally, individuals with dys-

TABLE 2
Phonics-based methods of reading instruction

Direct Instruction Reading (DSTAR)	Initially developed to teach economically disadvantaged students to read; emphasizes decoding skills with carefully sequenced daily lessons; instruction occurs in a small group, face-to-face with the teacher
Orton-Gillingham approach	Developed for students who have dyslexia; combines visual, auditory, and kinesthetic senses; reduces the English language to forty-four basic sounds and twenty-six letters; smaller units (such as syllables) are combined into more complex words
Spalding method (<i>The Writing Road to Reading</i>)	Developed as a method of classroom teaching and influenced by the Orton-Gillingham approach; teaches the sounds of letters, not their names, and pairs saying the sounds with writing them; previously learned sounds are combined into words that are familiar to the child

lexia must learn about their disability, be able to explain it to others, and request needed environmental assistance.

Examples of environmental assistance or coping techniques include using books on tape, taking oral instead of written examinations, participating in discussion groups, taking extra time to complete written or reading assignments, using study guides that give synopses of literature, and using word processor spell-check programs. Teaching methods that emphasize verbal instruction and hands-on experience in addition to peer readers or peer scribes also may be helpful.

WHAT HAPPENS WHEN CHILDREN WITH DYSLEXIA GROW UP?

Clinical experience and the popular press suggest a broad range of adult outcomes for people with dyslexia. Some become famous and successful despite their disability. Cher, Winston Churchill, Bruce Jenner, Greg Louganis, Nelson Rockefeller, and Henry Winkler are included in this group. Some people, particularly women, learn to cope with dyslexia so that they no longer appear to have the condition when formally tested. Nevertheless, a careful history and interview reveals that they still have many of the correlated characteristics. Other people can become successful professionals because of partners, spouses, and administrative assistants who read for them. Scientific studies are unclear on the proportion of people with dyslexia who actually have these successful outcomes, however.

Only a few studies have looked at the adult outcome of children with learning disabilities and dyslexia; many are flawed so it is difficult to interpret their results. For example, some studies followed individuals for an insufficient period of time. Others looked only at children from upper-middle-class families who attended private boarding schools for children with learning disabilities. Some failed to include an adequate comparison group of nondisabled individuals or included mentally retarded individuals among those with learning disabilities.

Among all the outcome studies, perhaps the most complete and least flawed is reported by Spreen (1987). This study followed West Coast Canadian children from the ages of eight to twelve years into their mid- to late-twenties. Children with mental retarda-

tion were excluded and a group of nondisabled children were followed so that meaningful comparisons could be made. The study was naturalistic, meaning the treatment or intervention each child received was allowed to vary naturally based on individual circumstances, such as family and school resources. Adult adjustment factors, such as education completed, level of employment, social and emotional adjustment, and delinquency were examined. Individuals' parents also were followed, in order to obtain their perspective on their children's adult adjustment.

The adult outcome of children with dyslexia was not analyzed separately from other types of learning disabilities, but the results are discussed here because of the study's relatively good quality. Not surprisingly, Spreen and his colleagues found that the adult adjustment of individuals with learning disabilities was best among those raised in families whose parents completed more years of education and were employed at higher levels. This result is not surprising because of the economic and community resources available to more affluent families. These resources can assist with remedial tutoring, education, and job search and placement for a child with a learning disability.

Adults who had learning disabilities as children had less favorable outcomes than their nondisabled peers, when results of the study were considered as a group. This finding held, even when disabled and nondisabled people were matched for intellectual ability. Individuals who had learning disabilities as children stopped their educations sooner, were employed at lower level jobs, earned lower salaries, and received mental health services more often than their nondisabled peers. They did not outgrow their learning problems, and in fact, measures of verbal intelligence and academic achievement at 8 to 12 years of age strongly predicted adult outcome ten to fifteen years later. Spreen (1978) summarized the study's results in this way:

The results have shown considerable social and economic implications: not only do these youngsters suffer through a miserable and usually shortened school career, live a discouraging social life, full of disappointments and failures, they also have fewer chances for adequate employment and advanced training. [p. 133]

This article's readers should recognize that these conclusions are summarized for the 226 individuals

with learning disabilities who agreed to participate in the study and that many participants had learning problems that affected more than just reading and spelling. As such these results do not adequately predict what the adult outcome might be for an individual case. Although some of the study participants were underemployed, imprisoned, or living in institutional settings, still others owned their own homes and lived comfortably within their communities. Additionally, individuals who did not agree to participate were not studied and we do not know whether they differed from study participants. For example, participants may have been less well adjusted than nonparticipants, hoping that the study would benefit their circumstances. On the other hand, participants may have been better adjusted, hoping that participation would demonstrate their success.

Clearly these results address the need for early identification and intervention with individuals who have dyslexia or other learning disabilities. Dr. Pennington suggests that intervention should begin early, also, with an emphasis on skills in phoneme awareness and a phonics approach to reading. The study by Spreen (1987) and colleagues noted a positive association between family socioeconomic status, private tutoring in childhood, and later ability to hold a permanent job. Further confirmation of this association is apparent in studies of more restricted groups of children with learning disabilities. Specifically, students from middle- or uppermiddle socioeconomic levels who attended private schools for the learning disabled or received substantial, individual remedial help, as a group, were more frequently employed and employed in higher-status occupations than was the group of learning disabled individuals from Spreen's study.

In addition to the needs for early identification and intervention with phoneme awareness skills and phonics-based reading programs, numerous studies have shown that one-to-one tutoring is more effective than small-group or classroom-based interventions. Individual tutoring probably contributes to vocational outcome by improving academic skills and probably re-

flects the individual's persistence and positive parental involvement. In summary, individuals with dyslexia, who have received individualized, phonics-based tutoring, should have good adult outcomes, particularly if employed in situations that do not require much reading, such as human-resource management. Several helpful resources are listed below:

The International Reading Association, 800 Barksdale Road, Box 8139, Newark, DE 19712-8139. This association publishes several brochures that help parents develop their child's reading skills.

The Orton Dyslexia Society, 724 York Road, Baltimore, MD 21204. This society supports research and public awareness activities specific to dyslexia and other learning disabilities. They publish brochures and have yearly conferences.

ACLD (The Association for Children and Adults with Learning Disabilities), 4156 Library Road, Pittsburgh, PA 15234. This association supports public awareness and advocacy activities for individuals who are learning disabled.

Local public libraries and public school departments of special education can provide information on how to borrow books on tape.

BIBLIOGRAPHY

- AARON, P. G., & JOSHI, R. M. (1992). *Reading problems: Consultation and remediation*. New York: Guilford Press.
- PENNINGTON, B. F. (1991). Dyslexia and other developmental language disorders. In *Diagnosing learning disorders: A neuropsychological framework* (pp. 45-81). New York: Guilford Press.
- ROONEY, K. J. (1991). Controversial therapies: A review and critique. *Intervention in School and Clinic*, 26, 134-142.
- SPALDING, R. B., & SPALDING, W. T. (1969). *The writing road to reading*. New York: Quill/William Morrow.
- SPREEN, O. (1987). *Learning disabled children growing up: A follow-up into adulthood*. Lisse, Netherlands: Swets & Zeitlinger.

INGRID N. LECKLITER

E

EASTERN VIEWS OF INTELLIGENCE It is difficult to know when Eastern views became distinct from Western ones; no definite point can be identified. Nevertheless, history provides a context for our understanding of how the distinctions began to emerge. The Eurasian continent, the Middle East, and Africa are inhabited by peoples who found ways to travel, trade, and contact one another throughout pre-history. By about 2000 B.C. the movement of nomadic Aryan tribes from Central Asia to the south and east, into India by way of Persia (present-day Iran), and in the other direction, to the west and north, to Greece, Rome, Scandinavia, and Ireland, resulted in some shared mythologies, philosophies, and the Indo-European roots of Greek, Latin, and Sanskrit. The Greeks, Romans, and Near Eastern peoples, from about 600 B.C. to about 600 A.D., established eastern outposts in India and, by means of the Silk Road and Indian Ocean, to China.

East–West contacts were not extensive, however, so basic cultural traditions, religions, and philosophies developed in relative isolation. Judeo-Christian philosophy shaped much of the Near East and Europe, beginning with the later Roman/Byzantine Empire; Hindu and Buddhist philosophies shaped much of the Indian subcontinent, Southeast Asia, China, Japan, and Korea. By the seventh century, when Islam was spreading both east and west from Arabia, Muslim conquerors faced resistance from the Jews and Chris-

tians of Europe as well as from the Hindus and Buddhists of Asia.

The study of human intelligence is recent—since the late nineteenth century—and it is secular, objective, and quantitative. Eastern views of human intelligence, therefore, are heavily influenced by the research and writings of psychologists in Europe and the United States, but the Eastern views of human intelligence remain grounded in Eastern tradition.

INTELLIGENCE AND ITS RELATION TO KNOWLEDGE

To understand the Eastern concept of intelligence, we must first define what it is. The meaning of intelligence in Eastern philosophies—that is, Hindu and Buddhist philosophies—is closely associated with moral and religious attitudes. Therefore, it is difficult to separate a person's intelligence from moral and religious values expressed in behavior. The most commonly used synonym for intelligence is *Buddhi* (Zimmer, 1951). It means waking up, noticing, recognizing, understanding, and comprehending. In contrast to the usual meaning of intelligence, as understood in Western literature, *Buddhi* includes such things as determination, mental effort, and even feelings and opinions in addition to such intellectual processes as knowledge, discrimination, and decision making.

In Eastern views, the best use of intelligence is in acquiring knowledge; however, it is “true knowledge” that is being sought. Intelligence in its highest form is focused on actions for obtaining true knowledge. But what is pure intelligence or true knowledge? How does one focus intelligence or apply it?

First, the basic mechanisms of gathering knowledge are considered. The Eastern views here do not differ very much from contemporary Western views. The individual is active through five sense organs (perception) and five motor organs (action). These ten external organs are the means by which the individual experiences the external world. The five motor organs are for speech, grasping, locomotion, reproduction, and excretion. It is not usual in contemporary Western psychology to pay particular attention to the motor organs while discussing intelligence, but, like the sensory organs, they contribute to our experience and knowledge.

The ten external organs are contrasted to the three so-called internal organs, that is, the mental functions. These concepts cannot be precisely translated into English, but the three internal organs are best described as the mind, which both reasons and is the activity of reasoning; the ego, or ego-sense, which both “ego-senses” (described below) and is the activity of “ego-sensing”; and intelligence (Buddhi), which both is and attains true knowledge.

The mind collaborates with the ten external organs in collecting experiences; its function is thinking, especially making inferences. Next, experiences are presented to the ego, or ego-sense, which tries to assimilate the information, and then presents it for evaluation and discrimination by intelligence, or Buddhi. The mind is therefore the first level of psychological contact with sensory experience; its function is to make inferences, just as the functions of the external organs are to see, hear, touch, and so on. The ego, or ego-sense, makes the mind’s representations one’s own—in other words, its function is to put the stamp of the individual on the representations presented by the mind. Both representations and evaluations are goal driven, the goal being the understanding of the true nature of things, or the obtaining of pure knowledge.

This goal is facilitated by certain attitudes of the individual. One is the basic motivation for pursuing

the acquisition of knowledge, even when such pursuit is hard to do. Attitudes promoting understanding aid us in distancing or detaching ourselves, so that we may promote freedom from the worries and anxieties of life while we are trying to understand the true nature of experience.

This kind of intelligence is “discerning,” or “pure,” intelligence, which is superior to the ordinary kind of intelligence. The latter has no single direction; it may be focused on several unrelated ideas or on understanding many discordant objects and desires in one’s life and the world. This kind of ordinary intelligence can give us false knowledge because it is excessively influenced by the present experience and transitory desires of an individual, rather than being directed toward the higher goal of acquiring true knowledge.

THE ELUSIVE PURE INTELLIGENCE

Intelligence in its purest form may be quite elusive; however, it can be attained. As mentioned above, some detachment and distancing are necessary so that intelligent activity may be free from the pressure of immediate sensory experience and memory before we have a chance to sort them out and evaluate them as helpful or unhelpful for obtaining true knowledge. Impure intelligence, therefore, does not have or is not guided by judgment and discrimination. Rather, it is a victim not only of the pressures of immediate experience and memories but also of habitual associations, desires, prejudices, and prejudgments.

It may sound strange to someone educated in the West, but the goal of pure and discriminating intelligence is to realize that the knower and the thing known are not different and in fact are not even two separate entities. The distinction between the two may be a necessary stage for the development of intelligence. That stage has to be crossed, however, and ego-centric thoughts, such as, “It is I whose experience is unique to me,” “It is I whose memory is unique to me,” have to be ultimately given up to obtain true knowledge by exercising discerning intelligence. Thus, by definition, the true representation of knowledge cannot be egoistic. We continue to remove our own biases, prejudices, and prejudgments, which create barriers between us and what we want to know. We have created the idea of an imaginary someone who is

residing within our minds and wants to know. We seem to need this “little person” inside our head who, as it were, perceives, memorizes, and thinks. Such an imaginary person in the head is called a “homunculus” in contemporary psychology. This problem is recognized in Eastern views. The necessity to have a duality between an agent who knows, on the one hand, and objects or ideas that are known, on the other, has been recognized as the consequence of egoistic thinking. A homunculus is not necessary for another reason if the Buddhist philosophy is considered: Intelligence is an activity of the mind, and since the mind’s activities follow one another in a stream of consciousness, lasting only for a moment, a sense of continuity in mental life can emerge. This idea of a stream of consciousness is similar to that of William James (1890).

EVOLUTION OF INTELLIGENCE: TOP-DOWN OR BOTTOM-UP?

A nagging question in contemporary psychology relates to consciousness. Is it an outcome of brain processes? If it is, can it also somehow influence the processes of the brain? We cannot solve the question of the relationship between consciousness and the physical structure of the brain here, but we can describe Eastern views. According to these views, at the top is pure consciousness, which does not have a material basis. First to evolve from pure consciousness is discriminative intelligence, or *Buddhi*. Evolving out of this discriminative and evaluative intelligence is ego-sense, or consciousness of one’s own identity as an individual. The mental functions evolve secondarily out of intelligence and ego-sense. These mental functions are reflected in the individual’s ability for perception and actions. Next to evolve are the five perceptual abilities and the five motor abilities. These abilities then are attached to five basic elements, or natural substances: earth, light, water, wind, and space. Although this order of evolution is apparently contrary to common sense, it can be justified as a subjective experience. Consider whether the top-down view can be accepted: With the help of discriminative intelligence, the materials that make up sensations, emotions, memories, and impulses for actions are observed. The elements of nature assume their form only

in our subjective world, and only through the so-called internal organs (mind, ego-sense, and intelligence).

It is clear, then, that the world exists as represented in consciousness. This representation must be kept pure when, through the use of intelligence, the goal of attaining true knowledge is achieved. In Eastern views, the top-down evolution of intelligence and sensory experience makes it unnecessary to distinguish between a material world that is unconscious and a human world that is conscious.

INTELLIGENCE AND TEMPERAMENT

The representation of objects in consciousness follows a top-down process, as described above. Intelligence is used to reach as pure a representation as possible. Two specific obstructions block the use of intelligence, however. The first is desire, and the second is the inappropriate temperaments that constrain the use of intelligence. Desire is broadly conceived of as affect. The purpose of intelligent discrimination is foiled by affective overtones, which are very much like dirt that covers a mirror. Desire adversely influences the gathering of information through the sense organs, as well as the inferencing and synthesizing that are carried out by discerning intelligence. Neither positive nor negative emotions facilitate the acquisition of pure knowledge through the exercise of intelligence.

Besides affect, or desire, three kinds of temperaments may help or hinder the work of intelligence (Radhakrishnan, 1948). The helpful temperament is a spiritual temperament that fosters nonattachment, distancing its owner from affect, memory, and the stress of sensory experience itself. A second, and less helpful, temperament is an action-oriented one characterized by passion, attachment, and thoughts of rewards during the performance of certain actions. Such a temperament does not promote an unbiased gathering of knowledge and the exercise of discerning intelligence. The third, the worst temperament, the so-called dark temperament, is full of negative emotions and, in fact, promotes laziness and insensitivity. If intelligence is an awakened state, the dark temperament gives rise to its opposite. Every individual, however, possesses all three temperaments; one of these may be given preeminence at the time of using intelligence. Therefore, it is desirable to free intelligence from the influence of both the

action-oriented and the dark, slothful state and to allow the spiritual temperament to prevail. There are procedures discussed in the Eastern philosophies for promoting the spiritual temperament and suppressing the influence of the two others.

IMPROVING INTELLIGENCE

It is not easy to give up egocentricity and achieve a state of intelligence that leads to true knowledge. Anyone can attempt to do so, however. The usual method recommended is reflection. How does one reflect on egocentricity to free intelligence from its constraints? The following are some of the procedures that traditionally have been recommended (Goddard, 1952). Through ignorance, we lose a sense of unity and instead think of the body—our sense organs, feelings, and cravings. We must be determined to remove such ignorance, which makes us lose the sense of unity. Speech should not reflect prejudice, anger, or selfishness. Careless speech should be avoided. People should seek to engage in charity, goodwill, patience, and perseverance, as well as unity of purpose. They need to develop “right mindfulness,” that is, a habit of looking at the real meaning and significance of things without being distracted by their apparent similarities to and differences from other things. Finally, people need to develop “right concentration” by avoiding distractions and passing thoughts. All of these procedures require long practice so that they may become an individual’s second nature. Only then is it possible to have a stage of intelligence that allows attainment of real knowledge.

THE EAST AND THE WEST

Eastern and Western views of intelligence, despite a number of similarities, display major differences. First, the similarities—Intelligence operates on experiential knowledge and accumulated memory that have been organized and stored. Experiential knowledge is derived from perceptual analyses, inference, analogical reasoning, and materials learned from books and authorities. The mind is “in charge” of experiential knowledge as well as memory. Both views agree on this.

Consider, then, how the Eastern views are distinct from contemporary Western concepts of intelligence:

The activities of the mind are close to sources of knowledge and thus are likely to be contaminated, constrained, fickle, and idiosyncratic. Therefore, a discriminating intelligence is needed—distancing itself, witnessing the activities of the mind, and attempting to attain a real, or true, knowledge. Obstructions arise on the way to exercising discriminating intelligence, mainly in the form of emotions and unwanted temperamental characteristics. Desire, passion, and biases constrain intelligence. So do the inappropriate temperaments that orient an individual to the “darkness” of ignorance, sloth, and impulsivity or to “action” characterized by a desire for knowledge that is driven by immediate utility, pride, assertiveness, and so on. Western psychology does not consider that affect and right temperament are a prerequisite for the use of intelligence (see, e.g., Sternberg, 1987).

The other major difference is that the Eastern view regards intelligence as constantly evolving, getting closer to a state of “pure” knowledge by dint of an individual’s efforts. These efforts are not only directed toward gathering valid and appropriate knowledge, which Western views would support, but are also directed at the attempt to rise above the influences of desire and unfavorable temperaments through right ideas, effort, and reflections. Every individual is capable of this evolution. Since Eastern views reject the Western dichotomy between mind and body, they reject any sharp distinction between humans and other animals—all of them can be placed in hierarchical states of evolution of consciousness. In the case of humans, higher and higher states can be reached until a stage where the distinction between the knower and the known becomes nonexistent. At this point, intelligence becomes identical with knowledge, which is the culmination of the process of knowing.

How do the Eastern and Western views influence people’s behavior? The answer is a speculative one until careful ethnographic studies are carried out. Consider how the Easterner may attribute success, for example, to intelligence, effort, luck, and the possible order of these factors. Effort would be the prominent factor, along with luck, unless the person has an intellectual defect. Everyone can try to develop a higher stage of intelligence. Right attitudes and habits are critical for achievement. Right attitudes include appropriate moral and ethical values as well as right

temperament (a slothful temperament is totally inappropriate for achievement). But chance factors play an important role as well; among them are birth in an advantageous family, social environment, and chance meetings with a good mentor. Whether these beliefs have influenced the current achievement of Oriental children in schools, not only in China, Japan, and India but even in the United States, is something to consider.

BIBLIOGRAPHY

- GODDARD, D. (Ed.). (1952). *A Buddhist Bible*. New York: Dutton.
- JAMES, W. (1890). *The principles of psychology* (Vols. 1 & 2). New York: Holt.
- RADHAKRISHNAN, S. (1948). *Bhagavad Gita*. London: Allen & Unwin.
- STERNBERG, R. J. (1987). Intelligence. In R. Gregory (Ed.), *The Oxford companion to the mind* (pp. 375–379). Oxford: Oxford University Press.
- TOYNBEE, A. (1976). *Mankind and mother Earth*. New York: Oxford University Press.
- ZIMMER, H. (1951). *Philosophies of India*. London: Routledge & Kegan Paul.

J. P. DAS

EEG EVOKED POTENTIALS Brain activity can be demonstrated by graphing brain waves as measured by the electroencephalograph (EEG), an apparatus that senses and records. The electrical activity of the brain can be recorded from the scalp by placing electrodes on it in standard positions; the resulting electric potentials are channeled through the encephalograph, which converts the impulses into the vertical movement of a pen over paper.

The electrical activity of the brain was first demonstrated by the German psychiatrist Hans Berger in 1929, when he identified alpha waves. Most normal adults have an alpha rhythm (8 to 12 cycles per sec.), associated with wakeful relaxation, which largely vanishes with thinking or opening the eyes. Beta rhythm waves (13 to 30 cycles per sec.) are fast, and are associated with normal conscious waking conditions; they are often found in anxious people and can also be induced by drugs. Theta waves (4 to 9 cycles per sec.)

are slower than alpha waves and are found in many normal adults. EEGs vary according to age, emotions, metabolic changes, drugs, and states of consciousness—including sleep. They are used to monitor brain and metabolic diseases. They are also being used to study human intelligence.

In psychology, recent interest has focused on the psychophysiological variables correlated with intelligence—and presumably underlying it. Psychophysiological, hormonal, or other biological structures and functions most probably intervene between the most basic molecular and genetic structures of the human organism, such as deoxyribonucleic acid (DNA) and behavior; DNA does not influence behavior directly. It would be advantageous to find such intermediary biological processes, and it is expected that they would be somewhat influenced by societal environmental factors—an important consideration for childrearing.

It should be noted that investigations of the biological determinants of the intelligence quotient (IQ) have been undertaken largely to understand intelligence for scientific purposes, not necessarily for practical purposes (Eysenck, 1982); biological studies of the type described here might help make our theories more inclusive.

From the beginning of EEG studies in the 1930s, it was assumed that ordinary EEG measures might give some indication of brain functioning, but early results, mainly reported in the 1960s and 1970s, were contradictory and on the whole disappointing (see the review by Deary & Caryl, 1992). Recent studies, such as the work of Gasser et al. (1983), have been more encouraging: They found that greater power (more cycles per second recorded) within a given band—particularly the theta band—correlates positively with higher IQ (but more so in mentally retarded subjects than in normal subjects).

Correlations as high as .60 for retarded and .34 for normal groups have been obtained with a children's intelligence test, the Wechsler Intelligence Scale for Children. Coherence measures may also be important; Thatcher et al. (1983) found that bright children had lower coherence (i.e., similar patterns between different electrode sites), indicating that more *differentiated* brain activity is associated with intelligence. The most extensive study of spontaneous EEG and intelligence is by Giannitrapani (1985); an important conclusion

of his work is that brain activity, in particular, narrow electrical brain wave frequency bands, may be important for the correlation of an EEG measure with IQ. Consideration of different topographic sites is also important.

Our own study of ninety-three adult subjects (A. E. Hendrickson, 1982; D. E. Hendrickson 1982; see Eysenck, 1973) yielded results that support Ertl's findings. Table 1 gives the latency and amplitude correlations when auditory stimuli were used; combining latency and amplitude, and correcting for attenuation (the shrinkage associated with measures less than perfectly reliable) would suggest a correlation of between .6 and .7. Note that *short* latencies and *large* amplitudes predict *high* IQ, as one might have predicted. Other studies (e.g., Calloway, 1973; Gucker, 1973; Shucard & Horn, 1972, 1973) gave similar results, with the added suggestion that the amount of *variability* in the brain wave evoked by the presentation of a visual stimulus might correlate negatively with IQ. (Because of a poor signal-to-noise ratio in AEP work, large numbers of repetitions are required to get an *averaged* score for each point in the trace, very much as in reaction time measurement; the more variable the resulting brain wave scores are from one of the same stimulus repetition to another, the greater the measured variability of the EEG for that individual.)

Figure 1 shows the differences in AEPs for bright (high-IQ) and dull (low-IQ) children from Ertl's work. It will be clear that far more apparent than the differences in latency and amplitude are differences in the *complexity* of the wave form. Thus, the dull children have simple sinusoidal waves (a simple up-and-down wave), whereas the bright ones have a much more complex smaller set of waves superimposed on the simple sinusoidal ones. This suggests that the *variability* of the EEG wave response to the stimulus presented may be more important than the latency or amplitude of the brain's response. Specifically, modulations of the AEP waves can only be recorded if successive repetitions agree on their troughs and peaks; when there is variability in the form of the successive waves, then a trough on one evocation may fall on a peak on another, thus canceling out.

This notion, already latent in Ertl's work, was developed into a theory of intelligence and tested spe-

TABLE 1
Correlations of verbal, spatial, and total scores on the AH4 intelligence test with evoked potential latency and amplitude

	Verbal	Spatial	Total
Latency			
P_1	-.41	-.39	-.44
N_1	-.44	-.38	-.45
P_2	-.48	-.44	-.50
N_2	-.34	-.35	-.38
P_3	-.41	-.29	-.38
N_3	-.29	-.25	-.30
Amplitude			
A_3	.31	.10	.22
A_4	.45	.25	.37
A_5	.31	.19	.27

NOTE: Numerical subscripts refer to successive waves; P and N refer to positive and negative deviations, respectively.

cifically by A. E. Hendrickson (1982) and D. E. Hendrickson (1982). They argued that EEG variability is produced by errors in the transmission and processing of information through the cortex, probably at the synapses; that high IQ depends on relatively error-free transmission of information, and that for this reason IQ could be measured best in terms of variability. Hendrickson and Hendrickson used two measures of variability: one direct measure (variability at each 2-msec. point on the abscissa, averaged over all data points), and the indirect "string measure," so called because originally a string was laid over the wave form, and then measured for length, on the assumption that the complexity of the wave would increase the length of the wave. (Later, of course, more sophisticated methods were used.)

The Hendricksons tested 219 high school students and correlated the IQ scores of each (measured by the Wechsler Adult Intelligence Scale) against each student's EEG variability and string length, with resulting correlations falling around .70 and a composite EEG that correlated .83 with IQ, an unusually high correlation. In a second sample, consisting of sixteen court stenographers, EEG string length correlated .80 with IQ, EEG variability correlated -.66 with IQ. Con-

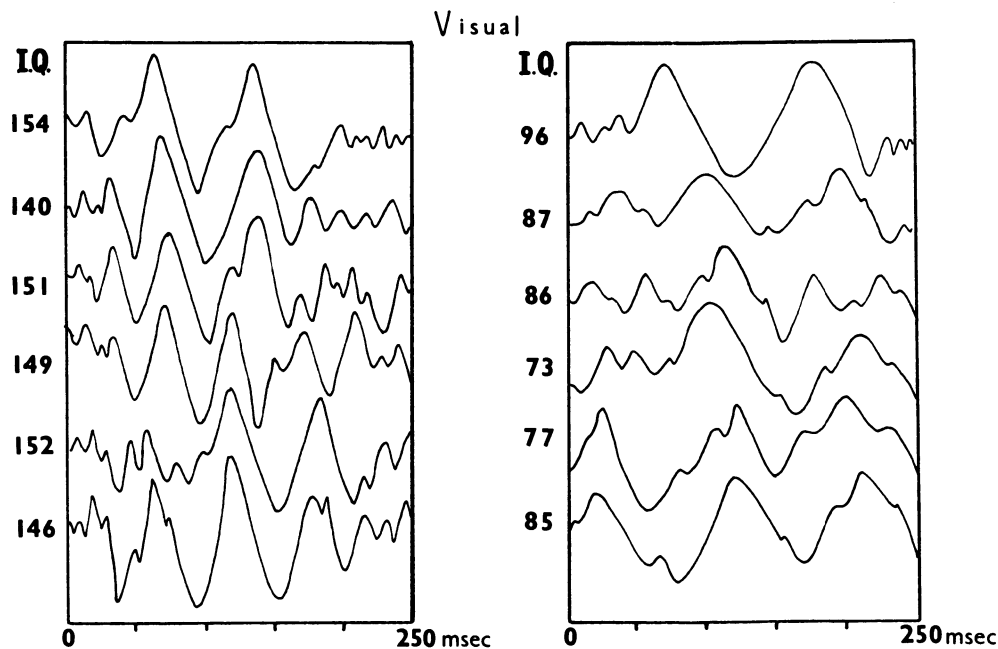


Figure 1
 Typical evoked potentials of bright (left) and dull children
 SOURCE: Eysenck, 1982.

currently, Blinkhorn and Hendrickson (1982) tested thirty-three undergraduates in a replication attempt and found a correlation of .54 between string length and IQ.

Later studies have sometimes confirmed and sometimes disconfirmed the Hendrickson and Hendrickson theory, although methodology, electrode placements, and other important details were usually sufficiently different to make comparisons difficult. Several studies suggest that the N140-P200 section of the EEG wave (i.e., the difference in height between the negative part of the wave at 140 msec. and the positive part of 200 msec.) is of particular importance in producing a true correlation between EEG and IQ, as well as giving the strongest heritability estimates. Clarification is clearly needed, but in terms of number of subjects, adequacy of methodology and clarity of effect, the original Hendrickson and Hendrickson studies remain unsurpassed (Eysenck, 1986), even though the size of the correlation is almost certainly too high to be replicable.

Schafer (1982, 1984) followed another line of argument, which might not be irreconcilable with the hypothesis of error-free transmission of information

through the brain. He argued that bright subjects should be characterized by *adaptability*, in the sense that they would commit fewer neurons than would dull subjects to processing known (repetitive) sensory input but would show greater response to unknown (novel) stimuli. For seventy-four normal adults he found correlations of around .60 between IQ and his Neural Adaptability Index. In a later study, Schafer (1984) tested forty-seven normal subjects and found a correlation of .59 between IQ and habituation, another measure of EEG adaptability. Correlation data in the original study were such that all subjects with low EEG adaptability had low IQs, but not all high-adaptability subjects had high IQs. Here, too, a repetition of that study following the exact format of Schafer, and using large groups of normal subjects (not students), would be required to be certain of the conclusions to be derived from this work.

A third approach, not unrelated to the EEG variability error theory, is based on the view that if the AEP responses of high-IQ individuals have more peaks and troughs in a shorter space of time, then analysis of the frequency spectra might provide a more formal

way of dissecting the underlying shape of the wave forms. In fact, Bennett (1968) and Weinberg (1969) found correlations of .40 to .60 between IQ and "the natural frequency of the dominant function" (Bennett) and IQ and mean spectral density (Weinberg).

In summary, we may draw the following conclusions. (1) Several EEG and AEP measures have been found to correlate with IQ, in adults as well as in children. These correlations range from about .20 or lower to .80, with or without correction for unreliability and range effects (i.e., the fact that when only students are used, the range of intelligence involved is less than it would be in an unrelated group, thus lowering all correlations with IQ). (2) The variables that an investigator must control in order to achieve a high correlation are not well understood; this makes successful replication of such studies difficult. Better understanding is needed of the precise meaning of AEPs at different frequency bands, different stimulus intensities, and different electrode placements. (3) Some findings do suggest theoretical interpretation, such as the discovery that differences relatively early in stimulus processing (around 200 msec. after the presentation of the stimulus) yield the best correlations with IQ. This may be interpreted to mean that we are dealing with stimulus-analysis mechanisms that may play a causal role in producing differences in ability. (4) Above all, as Deary and Caryl (1992) emphasize,

As well as continuing the search for the single EEG measure of brain activity which correlates most highly with g, it is now time for research workers to examine in greater detail the empirical relationships between IQ test scores and the numerous EEG and AEP measures currently available, with an emphasis on the information such brain wave measures provide about underlying mechanisms of intelligence.

(See also: EYSENCK, H. J.)

BIBLIOGRAPHY

- BENNETT, W. F. (1968). Human perception: A network theory approach. *Nature*, 220, 1147-1148.
- BLINKHORN, S. F., & HENDRICKSON, D. E. (1982). Averaged evoked responses and psychometric intelligence. *Nature*, 295, 596-597.
- CALLOWAY, E. (1973). Correlations between average evoked potential and measures of intelligence. *Archives of General Psychiatry*, 29, 553-558.
- CHALKE, F., & ERTL, J. (1965). Evoked potentials and intelligence. *Life Sciences*, 4, 1319-1322.
- DEARY, I. J., & CARYL, P. G. (1992). Intelligence, EEG and evoked potentials. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence*. New York: Ablex.
- ERTL, J., & SCHAFER, E. (1969). Brain response correlates of psychometric intelligence. *Nature*, 2123, 421-422.
- EYSENCK, H. J. (1973). *The measurement of intelligence*. Lancaster, PA: MTP Press.
- EYSENCK, H. J. (1979). *The structure and measurement of intelligence*. New York: Springer Verlag.
- EYSENCK, H. J. (ED.). (1982). *A model for intelligence*. New York: Springer.
- EYSENCK, H. J. (1986). The theory of intelligence and the psychophysiology of cognition. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3.). Hillsdale, NJ: Erlbaum.
- EYSENCK, H. J. (1992). The biological basis of intelligence. In P. A. Vernon (Ed.), *Biological approaches to the study of human intelligence*. New York: Ablex.
- EYSENCK, H. J., & BARRETT, P. (1985). Psychophysiology and the measurement of intelligence. In C. R. Reynolds & V. L. Wilson, (Eds.), *Methodological and statistical outcomes in the study of individual differences*. New York: Plenum.
- GASSER, T., VON LUCADON-MUELLER, I., VERLEGER, R., & BAECHER, P. (1983). Correlating EEG and IQ: A new look at an old problem using computerized EEG parameters. *Electroencephalography and Clinical Neurophysiology*, 55, 453-504.
- GIANNITRAPANI, D. (1985). *The electrophysiology of intellectual functions*. Basel: S. Karger.
- GUCKER, D. K. (1973). Correlating visual evoked potentials with psychometric intelligence, variations in technique. *Perceptual and Motor Skills*, 37, 189-190.
- HENDRICKSON, A. E. (1982). The biological basis of intelligence. Part I: Theory. In H. J. Eysenck (Ed.), *A model for intelligence*. New York: Springer Verlag.
- HENDRICKSON, D. E. (1982) The biological basis of intelligence. Part II: Measurement. In H. J. Eysenck (Ed.), *A model for intelligence*. New York: Springer Verlag.
- SCHAFER, E. P. W. (1982). Neural adaptability: A biological determinant of behavioral intelligence. *International Journal of Neuroscience*, 17, 183-191.

- SCHAFER, E. P. W. (1984). Habituation of evoked cortical potentials correlates with intelligence. *Psychophysiology*, 21, 597 (Abstract).
- SHUCARD, D., & HORN, J. (1972). Evoked cortical potentials and measurements of human abilities. *Journal of Comparative and Physiological Psychology*, 78, 59–68.
- SHUCARD, D., & HORN, J. (1973). Evoked potential amplitude change related to intelligence and arousal. *Psychophysiology*, 10, 445–452.
- STASSEN, H. H., LYKKEN, D. T., PROPPING, R., & BOMBEN, G. (1988). Genetic determination of human EEG: Survey of recent results on twins reared together and apart. *Human Genetics*, 80, 165–176.
- THATCHER, R. W., MCALASTER, R., LESTER, M. L., HORST, R. L., & CANTOR, D. S. (1983). Hemisphere EEG asymmetries related to cognitive functioning in children. In E. Verecman (Ed.), *Cognitive processing in the right hemisphere* (pp. 125–146). New York: Academic Press.
- WEINBERG, H. (1969). Correlation of frequency spectra of averaged visual evoked potentials with verbal intelligence. *Nature*, 224, 813–814.

HANS J. EYSENCK

ENVIRONMENT See NATURE, NURTURE, AND DEVELOPMENT.

ERROR OF MEASUREMENT Psychology has made many important contributions to the theory of measurement. Because psychological traits and characteristics cannot be directly observed, psychologists must rely on indirect measures of these constructs. Because all measurement in science is imperfect, psychologists have developed mathematical theories to assist them in determining how well tests measure psychological traits or characteristics.

The theoretical basis for mental measurement began in the late nineteenth century with attempts to measure intelligence (Spearman, 1904). From those beginnings, psychology has expanded its ability to accurately measure several other mental constructs. Today, the measurement theory derived from these early efforts is applied to all forms of testing, including achievement testing, licensure and certification exam-

inations, and clinical diagnoses, as well as measures of ability.

The advances that psychology has made were possible, in part, because of the conclusions from the theory of measurement error. Strictly speaking, psychologists and measurement theorists in other fields do not directly address the accuracy of measurement; rather, they approach the reliability of tests in terms of the relative absence of error. Classical test theory assumes that every measurement of an individual contains some amount of error. By making certain assumptions about these errors, for example, that they are as likely to be positive as negative, and that they are not systematically related to the actual observed test score, test theory derives a number of conclusions that define the properties of error in measurement. Although new approaches to the measurement of error have been developed, classical theory is still useful for understanding psychological measurement.

Psychological measurement is not more prone to error than measurement in other fields. All scientific measurement is subject to error. However, psychology is a self-conscious discipline, and the inferences about individuals that are often made on the basis of test scores increase the importance of a thorough understanding of the role of error in producing those scores (Nunnally, 1967, p. 173).

Classical test theory is based upon relatively simple assumptions about the nature of test scores. In this section, the assumptions that constitute the underpinning for measurement theory will be outlined briefly and their importance for the construction of instruments to measure mental concepts will be discussed.

TRUE SCORE THEORY

True score theory in psychology represents a measurement model that describes how the scores individuals receive on tests are derived. The “basic assumption” of true score theory (Gulliksen, 1987, p. 4) relates the observed scores (X) to the sum of two components: true score (T), and error (E):

$$X = T + E \quad (1)$$

This equation simply says that the two components must be added together in order for one to derive the observed score. Although other assumptions could be

made (for example, true scores and error scores could be multiplied together in order to derive the observed score), an additive relation is much simpler mathematically and allows the derivation of many other relationships that are useful in the construction of measurement instruments.

True score theory treats the observed score as a "fallible score" (Ghiselli, 1964, p. 221). The observed score is considered fallible because the theory assumes that the trait or characteristic measured by the test is measured with error. This error masks the true level of the trait or characteristic the individuals possess, the true score. True score theory describes how the true score and error score contribute to an individual's observed score.

The most important component of the observed score for an individual is the true score. One version of true score theory assumes that individuals possess stable characteristics that can be measured at a given point in time (Ghiselli, 1964, p. 221). This assumption does not necessarily imply that the characteristic for that individual is immutable over time. For example, the ability to solve problems in mathematics may increase as the result of study and instruction, or decrease because of lack of practice. This version of true score theory assumes only that a test can accurately measure how much mathematics knowledge an individual possesses when the test is administered. These stable characteristics are assumed to explain why individuals are able to correctly answer the items on the test. However, other versions of the theory derive the same results assuming distributions of scores over several testings of the same individual on the same test or on parallel tests (Allen & Yen, 1979, pp. 56–71; Nunnally, 1967, pp. 175–184). Many psychologists prefer this version because the existence of a platonic trait or characteristic need not be assumed (Lord & Novick, 1968, pp. 27–29).

The second component that contributes to the observed score an individual receives on a mental test is error. True score theory assumes that test scores fluctuate due to factors other than the individual's level on the characteristic that the test measures. These other factors are called *error* (Lord & Novick, 1968, p. 28). Two kinds of error may contribute to fluctuations in observed scores: *random error* and *constant error* (Gulliksen, 1987, p. 6). *Constant error* refers to fluctua-

tions in the observed scores that are systematically in the same direction. A tape measure that has been stretched from repeated use will always underestimate the length of the object measured, and consequently, the errors will always be negative. (The errors will always be negative because from the basic assumption, it follows that $E = X - T$, and therefore, the "true length" is longer than the tape measure shows.) Constant or systematic errors are not dealt with in true score theory, which is concerned primarily with random error.

If errors are random, by definition there will be no systematic pattern in the errors in observed scores: error is as likely to be negative as positive, as likely to be large as small. Since the errors display no systematic pattern, it is reasonable to assume that over a sufficiently large number of cases the average of the random errors will be 0. That is, true score theory assumes that in repeated testings of the same individual, there will be as many positive errors as negative ones, and that they will cancel out each other. Of course, for any given test and for any given individual, this assumption will not hold.

From the basic assumption and the fact that on average errors are equal to 0, it follows that the mean true score equals the mean observed score. Over a large number of tests, the average observed score an individual receives will equal that individual's true score. This is another definition of *true score* that does not require postulating the existence of an unobserved trait or characteristic of the individual.

If errors are random, it also follows that the error scores will not be related in any way to the true scores of a population of individuals. In other words, individuals with high true scores will be no more likely than individuals with low true scores to obtain positive or negative error scores. Further, the error scores on one test will not be systematically related to the true scores or the error scores on another test. These assumptions are extremely useful in the derivation of other characteristics of true scores and error scores.

PARALLEL TESTS

Another useful concept in true score theory is the idea of parallel tests, or parallel forms of the same test. Tests that measure exactly the same characteristic in

exactly the same way are *parallel*. In other words, it makes no difference which test is used in measuring an individual (Gulliksen, 1987, p. 11). Clearly, if one test is better than another for a particular purpose, the two tests are not parallel. The concept of parallel tests is important in estimating the reliability of mental measurements.

Mathematically, for two tests to be parallel, the true score that an individual receives on one test must equal the true score he or she receives on the other. Likewise, the errors of measurement on the two tests are equal. In terms of observed scores, the means and standard deviations (the square root of the variance and a measure of the variability of the scores) on parallel tests are equal, and the error scores on the two tests are not related in any way. Finally, psychologists can prove that the correlation of the scores on two (or more) parallel tests are equal. These results enable psychologists to compute estimates of the reliability of mental tests from the observed scores of individuals on a test.

RELIABILITY OF MENTAL MEASURES

The basic measure of RELIABILITY is the extent to which a test provides consistent results when it is repeated. Conceptually, reliability is the correlation between true scores and observed scores. In terms of true scores and error scores, reliability (ρ_{xx}') is the proportion of true score variance that an observed score explains. Typically, this is expressed as the ratio of true score variance (σ_T^2) to observed score variance (σ_X^2):

$$\rho_{xx}' = \sigma_T^2 / \sigma_X^2 \quad (2)$$

When that ratio is 1, there is no error variance and all the variance in the observed scores is due to true score; when that ratio is 0, the observed scores only reflect random fluctuations. Measurements include error so the values of the reliability coefficient range between 0 and 1.

The relationships among the variances of true, error, and observed scores that can be used to derive an equation for the error score variance that is expressible in terms of the variance in observed scores and the reliability of a test:

$$\sigma_E^2 = \sigma_X^2(1 - \rho_{xx}') \quad (3)$$

The square root of this equation has been given a special name, the *standard error of measurement* (σ_E), and can be computed from estimates of the standard deviation of the observed scores (s_X) on the test and the reliability of the test (r_{xx}'). This quantity is extremely useful in estimating the characteristics of the true scores for a test. There are several ways of estimating the reliability of a given test and the most familiar techniques will be discussed below.

ESTIMATES OF RELIABILITY

Because true score is a theoretical concept and cannot be observed directly, the actual correlation between these true scores and observed scores must be estimated from the observed scores on tests. The relation between the scores on parallel tests can be used to estimate the correlation between true scores and observed scores and to evaluate the accuracy of the estimate of true scores provided by a given test. In practice, it is not possible to construct or to administer large numbers of parallel forms of a measurement instrument. Therefore, psychologists have devised different methods for collecting information about an individual's performance on at least two tests or on the same test on different occasions that allow them to calculate estimates of reliability.

First, a measure of test-retest reliability can be computed by administering the same test to the same individuals on different occasions and correlating the scores from the two administrations. This correlation provides an estimate of the correlation between true scores and observed scores on the test. Since the same test is administered on both occasions, the assumption that the tests are parallel is met. However, two other factors may contaminate the results of this method. On one hand, individuals may be influenced in responding on the second administration by their performance on the first administration of the test. If individuals simply repeat answers they remember giving on the first testing, or if they benefit from the practice obtained by taking the test the first time, the results of the two administrations will not be independent. Therefore, the assumptions that allow psychologists to use the correlation between the observed scores on the two administrations as an estimate of reliability will not be met. On the other hand, the time

ETHNICITY, RACE, AND MEASURED INTELLIGENCE

be whatever it is that intelligence tests measure (Boring, 1923).

Despite or because of these problems in definition, different theories of intelligence have developed over the years that incorporate a cultural component. For example, R. B. CATTELL (1963) discusses two types of intelligence—fluid and crystallized. Fluid forms of intelligence are often viewed as nonverbal, culture free, and involving adaptive abilities. These abilities enable an individual to adjust to new situations and tasks. Fluid intelligence increases until adolescence and then levels off. Crystallized intelligence, on the other hand, may be based upon cultural transmission and assimilation, which is influenced by formal and informal educational factors throughout life. Cultural assimilation, in part, refers to the process of accepting the values and beliefs associated with the dominant culture often transmitted through the educational system. Crystallized intelligence is usually measured by traditional intelligence tests. (See FLUID AND CRYSTALLIZED INTELLIGENCE, THEORY OF.)

Arthur JENSEN (1973) proposed a two-level theory suggesting that abilities are of two “broad classes”—Level I and Level II. Level I abilities include those that involve short-term memory and rote learning. Level II abilities include those that involve mental manipulation, generalization, transfer of information, reasoning, and problem-solving. Level II abilities are associated with GENERAL INTELLIGENCE and may be genetically based. Although similarities exist between Cattell’s theory of fluid and crystallized abilities and Jensen’s Level I and Level II abilities, Jensen reports that fluid and crystallized abilities may be either Level I or Level II.

Jensen believes that these two levels of intelligence are present in all populations. Level I abilities are distributed evenly in all populations, and Level II abilities, on the other hand, are distributed differently in upper and lower social classes and in varying ethnic groups: “The majority of children now called ‘culturally disadvantaged’ show little or no deficiency in Level I ability but are about one standard deviation below the general population mean on tests of Level II ability” (Jensen, 1970, p. 25). Jensen and A. R. Inouye (1980) examined Level I and Level II abilities in relation to three ethnic groups—Asian Americans, blacks, and whites. Participants in this study included children in

and Asians obtained high scores on Level II measures (i.e. nonverbal IQ), and blacks obtained lower scores. Blacks and Asians also scored lower on Level I tests (i.e., memory tasks) in comparison with whites. Blacks differed more from whites on Level II abilities than on Level I abilities. Jensen and Inouye (1980) did not provide definitive interpretations of these findings. They cite the possibility of a genetic explanation because “differential selective pressures for different cognitive abilities in the evolutionary histories of these groups” (p. 49) may exist. In addition, cultural origins are also likely as differences are evident in “values, motivation, and styles of child rearing that currently predominate in each of these populations” (p. 49).

Other theories also suggest that various cultures may influence the development of one type of ability over another. For example J. P. Das (1973) noted that culture may create preferences for a particular form of information processing. R. Case and J. Pascual-Leone (1975) report that culture may influence a person’s mode of perception or cognitive style. Others indicate that race differences in intelligence cannot be attributed entirely to a single cause (Borkowsky, Krause, & Maxwell, 1985). They cite racial differences in strategy use in their research. Race-related differences in planning and other higher order intellectual processes are believed, in large part, to be related to early environmental stimulation in the home. R. J. Sternberg (1985) discusses a contextualist view of intelligence, suggesting that intelligence should be considered with respect to the circumstances in which a given individual develops and operates.

RELATIVE SCORING PATTERNS BY ETHNIC GROUP ON MEASURES OF INTELLIGENCE

Cultural differences between ethnic and racial groups influence the development of mental abilities, as cultures may foster certain unique ways of behaving. Ethnic and racial group differences have been of interest to researchers with respect to patterns of intellectual abilities (e.g., Backman, 1972; Jensen, 1973).

D. J. Reschly (1978) compared the performance of white, black, Chicano, and Native American Papago children on the Wechsler Intelligence Scale for Chil-

strument is appropriate as a measure of intellectual ability for different ethnic groups.

Studies have also used Jewish participants as a separate ethnic-religious group. T. Sowell (1978) noted an average IQ for this group of 109 on the California Test of Mental Maturity based upon data obtained in the 1940s. Current research also supports the finding that Jewish IQs tend to be above the national average. The higher IQs obtained may derive from the emphasis placed upon achievement and education within the Jewish culture. M. E. Backman (1972) compared Jewish whites, non-Jewish whites, blacks, and Asians from a project known as TALENT. Ethnicity accounted for 9 percent of the difference associated with the pattern of the abilities (i.e., shape) across the various intelligence measures. Ethnicity accounted for 4 percent of the difference associated with level of the pattern (i.e., higher or lower performance across the different tests). Gender accounted for the largest proportion of the differences in abilities.

R. L. Taylor and S. B. Richards (1991) compared the patterns of ability of black, Hispanic, and white children of ages 6 to 11 on the WISC-R. The WISC-R consists of twelve subtests and is the most widely used individualized intelligence test in the United States. A description of these subtests is presented in Table 1. The findings indicate that black children performed better on verbal tasks, Hispanic children performed better on visual-spatial tasks, and white children performed better on tasks requiring abstract thinking and knowledge. The Hispanic children scored highest on three of the eleven subtests of the WISC-R: Picture Completion, Block Design, and Object Assembly subtests. The group of white children scored highest on the Information and Similarities subtests, and the group of black children scored highest on the vocabulary subtest. Taylor and Richards (1991) noted that greater differences existed between the scores obtained within the groups of black and Hispanic children, whereas the white children were "more consistent regarding their overall pattern of skills" (p. 8). Hispanic children scored higher on the average in nonverbal than in verbal areas within group, whereas the opposite was true for the black children.

A comprehensive review of thirty studies using the Wechsler scales with Native Americans included ninety-three groups and subgroups (Vraniak, 1993).

Results indicate that these samples of American Indians obtained an average verbal IQ of 83 and performance IQ of 100. The average difference between the verbal and performance abilities was 17 (see also NATIVE AMERICANS).

A study of 6,869 children of various ethnic groups (white, Hispanic, black, Native American, and Japanese) examined the profiles of abilities yielded on the WISC-R (Suzuki & Gutkin, 1993). The group of white children scored at approximately 10 across each of the subtests, yielding a full scale IQ of 100 as the average for the white children. The groups of Hispanic, black and Native American children scored below the white children in overall intelligence (full scale IQ). The Native American and Hispanic groups demonstrated relative strengths in visual-reasoning in comparison to verbal-reasoning abilities as noted by subtest differences. The sample of black children obtained a profile that was relatively flat and consistently below the white average. The sample of Japanese children showed the most dramatic profile, with overall strengths noted on the visual-reasoning and performance subtests in comparison with the verbal subtests, with the exception of arithmetic. In this study, whites obtained an overall IQ of 102 (Verbal IQ = 102, Performance IQ = 103), Hispanics 91 (Verbal IQ = 88, Performance IQ = 96), blacks 88 (Verbal IQ = 88, Performance IQ = 90), Native Americans 88 (Verbal IQ = 83, Performance IQ = 95), and Japanese 106 (Verbal IQ = 96, Performance IQ = 117). These results must be interpreted with caution as within-ethnic group differences appeared among Native American tribal groups and possibly among the Hispanic subgroups. Regional differences were also observed among the samples of black children. Specifically, children from different tribal groups and from different Hispanic groups scored differently across the various subtests. In addition, the sample of Japanese children was an international group and may not have been representative of average Japanese Americans.

POSSIBLE REASONS FOR DIFFERENT ETHNIC SCORING PATTERNS

The examination of intelligence across different ethnic and racial groups becomes complex because a number of other variables can affect the measurable

TABLE 1
Description of WISC-R subtests

<i>Subtest</i>	<i>Nature of Items</i>	<i>Timed</i>	<i>Verbal/Performance Scale</i>
Information	This subtest consists of questions tapping the child's general fund of knowledge.		Verbal
Similarities	This subtest consists of questions requiring the child to identify similarities between pairs of words presented orally.		Verbal
Arithmetic	Arithmetical problems presented orally and in written form, requiring a verbal solution comprise this subtest.		Verbal
Vocabulary	This subtest asks the child to define words orally.		Verbal
Comprehension	Questions requiring social judgment comprise this subtest.		Verbal
Digit Span	In this subtest, the examiner verbally presents a series of random numbers and the child is asked to repeat the numbers in the same or backward sequence.		Verbal
Picture Completion	On this subtest, the child is asked to identify a missing part in a series of pictures.	✓	Performance
Picture Arrangement	The child is presented with a scrambled series of picture cards and is asked to sequence them to convey a story that makes sense.	✓	Performance
Block Design	A design consisting of colored blocks is illustrated with blocks or a picture. The child's task is to reproduce the design utilizing the blocks provided.	✓	Performance
Object Assembly	On this subtest, the child is asked to assemble puzzle pieces to form a recognizable whole as quickly as possible.	✓	Performance
Coding	The child is asked on this subtest to reproduce symbols in a code.	✓	Performance
Mazes	On this subtest, the child is asked to complete a series of paper/pencil mazes.	✓	Performance

SOURCE: Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children-Revised*. New York: The Psychological Corporation.

elements of intelligence. Research has indicated that there are intragroup differences in IQ. For example, Native American tribal groups and various Hispanic groups (Cuban, Chicano, Mexican) score differently from one another on measures of intelligence. Other issues include possible cultural bias and cultural loading on measures of intelligence, acculturation, assimilation, and socioeconomic status.

Cultural Bias. The study of intelligence and related abilities has been controversial, given the alleged bias in the measurement of mental abilities and the utilization of test scores. Court cases have challenged the use of intelligence tests with minority children; some black psychologists have demanded a moratorium on intelligence testing (Jackson, 1975); and attempts to reduce the disproportionate numbers of minority children in classes for the mentally retarded are examples of the incidents that have arisen in attempts to resolve the race, ethnicity, and IQ controversy.

Specifically in the 1960s, a condition known as the *six-hour retarded child* was identified. This phrase refers to those children who were classified as *retarded* based upon standardized IQ measures but who nevertheless were able to live productive lives within their communities. The “six hours” referred to the time these children spent in school. When these cases finally wound up in litigation, lawyers argued successfully that minority children were categorized disproportionately as “retarded” even though they functioned adequately in their homes and communities.

Decades earlier, in 1958, Shuey reviewed approximately 72 studies analyzing a total of 36,000 black children and concluded that they tended to score one standard deviation (approximately 15 IQ points) below white children on standardized IQ measures. This discrepancy between blacks and whites has been identified as one of the best documented phenomena in testing today (Reynolds, 1982).

Despite this empirical finding (also obtained in Suzuki & Gutkin, 1993) the discrepancies between different ethnic groups on IQ tests have been viewed by some writers as evidence that the tests are biased against particular minority groups. Discrepancies among scores earned by whites, blacks, and Hispanics often appear in the literature (e.g., Dean, 1979; Laosa, 1984; Reynolds, 1982, as well as studies cited previously).

Despite the views of supporters of the minority position, they have often been unable to prove conclusively that many of our most commonly used measures of intelligence are biased against particular minority groups (e.g., Oakland & Parmelee, 1985). There is a difference between cultural test bias and test fairness. Test bias is defined as “systematic error in the estimation of some ‘true’ value for a group of individuals while test fairness relates to the application of a test in some decision-making process such as selection,” for example, special education versus regular classroom placement (Reynolds, 1982, p. 186).

Cultural Loading. Intelligence tests can be culturally loaded without being culturally biased. *Cultural loading* refers to the degree of cultural specificity present in a particular test. American intelligence tests include items that refer to the history of the United States from a majority perspective. Virtually all tests are bound in some way to the unique aspects of the culture in which they were developed (Reynolds, 1982). Tests must to some extent be relevant to the definition of intellectual behavior within a particular culture.

Other Factors. Researchers have observed discrepancies in intelligence based on acculturation and assimilation. Groups of individuals of different national origins who have been in the United States longer appear to score differently from those groups who have recently immigrated (e.g., British and European groups versus Haitians). Regional or geographic differences (e.g., the southern versus the northern states of the United States) have been noted in the literature with respect to IQ discrepancies as well. Language differences and cultural values have also been factors influencing measured intelligence. Culture-free or nonverbal tests (measuring perceptual and visual-reasoning skills) recur as alternatives to traditional IQ tests. Ethnic minority children often do not perform any better on “culture-fair tests,” however, than on traditional measures of intelligence (Sattler, 1988).

A number of specific problems involving the research designs and related methodologies have hampered progress in discerning the pattern of intellectual abilities for various ethnic groups. Results remain inconclusive and inconsistent because of mixed samples of referred versus nonreferred participants in the same study, small sample sizes, and unrepresentative groups.

When sample sizes are too small or include individuals with a mixture of other uncontrolled variables, such as handicapping conditions, the results of such research cannot be generalized to the ethnic group as a whole. Within the context of these limitations, the remainder of this essay addresses the comparison of different ethnic groups on various intelligence measures with respect to socioeconomic status, regional differences, and language.

In an early, frequently criticized study, G. S. Lesser, G. Fifer, and D. H. Clark (1965) provided one of the first demonstrations of the apparent differences in profiles of abilities among ethnic groups. The researchers administered a modified version of the Hunter College Aptitude Scales for Gifted Children (an instrument for assessing verbal ability, reasoning, numerical reasoning, and space conceptualization) to first-grade pupils from four ethnic groups in New York City—Chinese, Jewish, black, and Puerto Rican. The major findings were as follows: (1) differences in social class (e.g., poor, middle class, or high-income families) were associated with some significant differences in the level of each mental ability examined, but did not produce significant differences in the patterns of these abilities; (2) differences in ethnic background did correspond with significant differences in both the level of each intellectual ability measured and the patterns among these abilities; and (3) social class and ethnicity affected the absolute level of each mental ability but did not affect the patterns among these abilities. Examination of within-ethnic group profiles yielded the following: (1) Chinese children did better on space, number, and reasoning tests and relatively poorer on verbal tests; (2) black children obtained their highest scores in verbal ability and lower scores on number, reasoning, and space; (3) Jewish children did best on verbal tasks and performed lowest on spatial tasks; their numerical abilities were relatively high; (4) Puerto Rican children demonstrated less variability in their measured abilities in comparison with the other groups; their best performance occurred in tasks involving spatial abilities, and their poorest performance was on the verbal measure.

Similar overall findings were also reported by T. Sowell (1978) in his work with the Urban Institute. The research he conducted incorporated approximately 70,000 IQ records of students across the

United States from over a dozen ethnic groups. Data were collected over varying periods, some up to fifty years. Sowell examined the data to determine historical patterns of mental test results. His review of studies in the area of IQ differences among American ethnic minorities yielded increases as large as 20 IQ points over decades as “past disadvantaged groups rose socioeconomically” (p. 229). For example, Sowell noted that for the United States, the pattern of Asian IQ scores is “a relatively simple one of lower-than-average IQs in the early years (of the 20th century) and higher-than-average scores in the later years” (pp. 212–213). He also reported that Asian-American children scored higher than white children on particular subtests that did not involve a language component, suggesting that language may affect performance on traditional IQ measures. Sowell also found that the overall median IQ of black Americans ranged from 82 to 98 across the fifty-year span. Sowell observed regional differences for black children as discrepancies were evident between black children living in the South (whose IQs were lower) in comparison with other areas of the United States. Sowell reported that black Americans living in the South generally score lower than other black Americans, possibly due to environmental factors. For example, “poorer schooling and other social constraints on blacks in the South have led to poorer intellectual performances” (p. 228). He also noted that scores for blacks (on particular tests) have “risen by larger increments than have scores in the general population in response to such environmental improvements as better teaching or better test familiarity and test environment” (p. 229).

Sowell’s study also revealed within-ethnic-group differences as surveys conducted by the Urban Institute between 1930 and 1970 indicate that Chinese and Japanese-American median IQs ranged from 101 to 108, and Mexican-American median IQs ranged from 82 to 87. Similar score ranges were found for Puerto Ricans and Hispanics. A 10-point increase resulted for Mexican Americans on the California Test of Mental Maturity between 1950 and 1960, and a systematic rise in IQ for Puerto Rican children was found as the number of years of attendance in schools in the United States increased for different samples of Puerto Ricans.

The importance of language in relation to assessment is critical. Although tests of intelligence may be

translated into various languages, the different versions of such tests may not be parallel with respect to the abilities measured. Recall that the test items remain culturally loaded, that is, relevant to the culture in which they were developed. In addition, researchers must address the language dominance and language proficiency of the children being examined "because it provides information necessary to conduct accurate assessments of other intellectual abilities" (Olmedo, 1981, p. 1083).

In conjunction with lower socioeconomic status are a number of other factors that impact the measured abilities of particular ethnic groups. D. Amante and colleagues (1977) report that "levels of neurological integrity vary along a socioeconomic gradient and between ethnic groups" (p. 524). *Neurological integrity* refers to the level of intactness of the brain. In particular, they suggest that the impact of malnutrition/undernutrition, as well as inadequate prenatal and postnatal care, are specifically related to dysfunction in children and linked to the socioeconomic status of minority groups. However, many other studies relating to these same factors have not shown a relationship between intelligence test scores and malnutrition or any of these other variables.

CONCLUSIONS

Numerous investigations have provided group comparisons in the levels of measured intelligence of different ethnic populations. The findings of such studies have been criticized by writers who believe that they depend on the specific measures of ability used and the ethnic groups and samples included. Nevertheless, most studies have reported differences in the average score on tests of intelligence, as well as in patterns of ability across the various ethnic groups. These results should be interpreted with caution, as this same body of research also reveals wide ranges of ability within each ethnic group (e.g., regional and tribal differences). In addition, factors such as language, acculturation, cultural differences, socioeconomic status and other issues, may impact differences in measured intelligence associated with ethnicity. In conclusion, although disparities in measurable average levels of intelligence have consistently been reported, the reasons for these racial-ethnic differences, as well as for dif-

ferences across different samples within the same ethnic-racial groups, remain unclarified.

(See also: ASIAN AMERICANS; HISPANICS; JAPANESE; NATIVE AMERICANS; RACE AND IQ SCORES.)

BIBLIOGRAPHY

- AMANTE, D., VANHOUTEN, V. W., GRIEVE, J. H., BADER, C. A., & MARGULES, P. H. (1977). Neuropsychological deficit, ethnicity, and socioeconomic status. *Journal of Consulting and Clinical Psychology, 45*, 524-535.
- BACKMAN, M. E. (1972). Patterns of mental abilities: Ethnic, socioeconomic, and sex differences. *American Educational Research Journal, 9*, 1-12.
- BORING, E. G. (1923). Intelligence as the tests test it. *New Republic, 35-37*.
- BORKOWSKY, J. G., KRAUSE, A., & MAXWELL, S. E. (1985). On multiple determinants of racial difference in intelligence: A reply to Jensen. *Intelligence, 9*, 41-49.
- CASE, R., & PASCUAL-LEONE, J. (1975). Failure of conservation training of disadvantaged Black teenagers: A neo-Piagetian interpretation. *Perceptual Skills and Motor Development, 40*, 545-546.
- CATTELL, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology, 54*(1), 1-22.
- DAS, J. P. (1973). Structure of cognitive abilities: Evidence of simultaneous and successive processing. *Journal of Educational Psychology, 65*, 103-108.
- DEAN, R. S. (1979). Distinguished patterns for Mexican-American children on the WISC-R. *Journal of Clinical Psychology, 35*, 790-794.
- JACKSON, G. D. (1975). Another psychology view from the Association of Black Psychologists. *American Psychologist, 30*, 88-93.
- JENSEN, A. R. (1973). Level I and Level II abilities in three ethnic groups. *American Educational Research Journal, 10*(4), 263-276.
- JENSEN, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- JENSEN, A. R., & INOUIYE, A. R. (1980). Level I and Level II abilities in Asian, white and black children. *Intelligence, 4*, 41-49.
- LAOSA, L. M. (1984). Ethnic, socioeconomic, and home language influences upon early performance on measures of ability. *Journal of Educational Psychology, 76*(6), 1178-1198.

- LESSER, G. S., FIFER, G., & CLARK, D. H. (1965). Mental abilities of children from different social-class and cultural groups. *Monographs for the Society for Research in Child Development*, 30 (4, Whole No. 102).
- OAKLAND, T., & PARMELEE, R. (1985). Mental measurement of minority group children. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories of measurement and application* (pp. 699–736). New York: Wiley.
- OLMEDO, E. L. (1981). Testing linguistic minorities. *American Psychologist*, 36, 1078–1085.
- RESCHLY, D. J. (1978). WISC-R factor structures among Anglos, Blacks, Chicanos, and Native-American Papagos. *Journal of Consulting and Clinical Psychology*, 46, 417–422.
- REYNOLDS, C. R. (1982). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (pp. 178–208). New York: Wiley.
- SATTLER, J. M. (1988). *Assessment of children* (3rd ed.). San Diego, CA: Jerome M. Sattler.
- SHUEY, A. M. (1958). *The testing of Negro intelligence* (2nd ed.). New York: Social Science Press.
- SOWELL, T. (1978). Race and IQ reconsidered. In T. Sowell (Ed.), *Essays and data on American ethnic groups* (pp. 203–229). New York: The Urban Institute.
- STERNBERG, R. J. (1985). Cognitive approaches to intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 59–118). New York: Wiley.
- SUZUKI, L. A. (1992). *Ethnic ability patterns on the Wechsler Intelligence Scale for Children-Revised: Relationship to theories of intelligence and cognition*. Unpublished doctoral dissertation, University of Nebraska, Lincoln.
- SUZUKI, L. A., & GUTKIN, T. B. (1993). *Racial/ethnic ability patterns on the WISC-R and theories of intelligence*. Paper presented at the 101st American Psychological Association Convention, Toronto, Canada.
- TAYLOR, R. L., & RICHARDS, S. B. (1991). Patterns of intellectual differences of black, Hispanic and white children. *Psychology in the Schools*, 28, 5–8.
- VERNON, P. A., JACKSON, D. N., & MESSICK, S. (1988). Cultural influence on patterns of abilities in North America. In S. H. Irvine & J. W. Berry (Eds.), *Human abilities in cultural context*. New York: Cambridge University Press.
- VRANIAC, D. (1993). *Technical report to the Bureau of Indian Affairs* (BIA). Madison: University of Wisconsin, Madison.
- VROON, P. A. (1980). Intelligence on myths and measurement. In G. E. Stelmach (Ed.), *Advances in psychology*, 3 (pp. 27–44). New York: North-Holland.

LISA A. SUZUKI

DAMIAN A. VRANIAC

EVOKED POTENTIAL See EEG EVOKED POTENTIAL.

EVOLUTION OF HUMAN INTELLIGENCE

The study of the evolution of human intelligence provides a window into the history of what makes us as humans so special and distinct from all other species. Our intelligence alone sets us apart as qualitatively different from, and in many ways superior to, the other animals on our planet. The questions of why and how we are different intellectually have intrigued scientists from a number of different disciplines, including psychology, anthropology, biology, and evolutionary science. Taken together, the work of these scientists has provided a great deal of insight into how we differ from other species and suggests several plausible hypotheses concerning why these changes might have occurred.

How and why we have evolved intellectually are the topics of this entry. Ideas and suggestions from different scientists will be discussed below. As will become evident, none of the theories can be accepted in their entirety, but collectively they provide some interesting perspectives on our evolution and some thought-provoking ideas about how this evolution came to occur.

UNIQUENESS OF HUMAN INTELLIGENCE

The degree to which the human species is unique among animals is disputed among theologians and scientists. For instance, some religions (e.g., Christianity) view animals and humans as spiritually distinct, whereas others (e.g., Hinduism) view them as at different points on a continuum (Gibson, 1990). In the scientific realm, this argument has been spurred by evolutionary theory. Although heated disagreement

still exists between evolutionists and creationists, the data suggesting that we evolved from earlier life forms is convincing. As H. J. Jerison (1982, p. 737) stated, "The evidence is overwhelming." Yet, even among evolutionists there is no agreement on whether there is a qualitative or quantitative difference between humans and other species. The primary difference between the species, of course, centers on their respective intellectual capacities. The main question is whether human intellectual capacity differs from other species in terms of total brain size and memory and information-processing capabilities (e.g., Jerison, 1973, 1982) or in terms of total brain reorganization (e.g., Holloway, 1966, 1968). Support for, and theories favoring, each of these views can be found.

A working definition of intelligence is central to this discussion. Two types of intelligent behavior were identified by R. J. Sternberg and W. Salter (1982) as summarizing intellectual expressiveness: adaptive and goal-directed behaviors. Intelligent behavior that is adaptive is successful in meeting the challenges presented either by the individual or by the environment. These challenges and the behaviors they generate may differ across species, but the concept is useful because of its cross-species applicability. Adaptive behavior is not sufficient to demonstrate intelligence, however. Behavior must also be goal-directed to demonstrate intelligence. Again, the goals and the ways in which they are reached will differ across species and individuals.

BRAIN SIZE AND NEURAL REORGANIZATION

It has become clear that brain size is related to degree of intelligence across species, although this does not clearly hold within species, where small differences in brain size are not relevant. Yet, absolute brain size is not a useful measure. Rather, allometric brain size (brain size relative to body size) is the measure that relates to intelligence (Gibson, 1990). For instance, the brains of elephants and whales are larger than those of humans, but when adjusted for body size, human brains are allometrically the largest (and, we believe, the most intelligent).

A person living today has a brain almost four times as large as one of our human ancestors who lived more

than 3 million years ago. The rapid growth of the brain in *Homo* was tied only somewhat to increasing body size. In other words, the increase in brain size was not merely a by-product of natural selection for bigger bodies but also an unprecedented evolutionary event (Falk, 1992).

The increase did, in certain ways, follow general patterns for the evolution of larger brains that have occurred across a number of species: Neurons became larger and increasingly spread out, and the cerebral cortex became more and more convoluted to accommodate its increase in surface area. In the human brain, the increase in cortical and resulting convolutions occurred primarily in the prefrontal and posterior association areas. The prefrontal areas mediate goal-directed behavior and provide the formative bases for personality. The posterior association areas integrate input from the senses (hearing, smell, etc.) and put them together in meaningful ways.

The size of the brain does seem to be related to intelligence level. The relationship between increased brain size and higher general intelligence has held over examination of a wide range of species (Jerison, 1982). A number of researchers have also postulated that neural reorganization is a key component in the evolution of human intelligence (e.g., Gibson, 1990). For example, the human brain is three times as large as the brain of a chimpanzee, yet has only 1.25 times as many neurons. This relatively low density of neurons is accompanied by more interneuronal connections (axons and dendrites), suggesting that the human brain is designed for a high level of intercellular communication and integrated information processing (Parker & Gibson, 1979). The increase in connections for each neuron provides humans and other large-brained species with increased differentiation of sensory and motor units, therefore making a wider variety of sensory and motor behaviors available. This variety allows organisms to combine and recombine a vast array of sensory and motor behaviors, thus allowing them to exhibit greater intelligence (Gibson, 1990).

Thus, it is clear, as described above, that our neural interconnections, and consequently brain size, have increased dramatically in complexity and size in the last few million years. Given these changes, the pressing question of interest now is how and why this evolution

has taken place. Several theories have been proposed, but none are completely satisfactory. A brief summary of these theories and possible problems with each are presented below.

ONTOGENY RECAPITULATES PHYLOGENY

The increasingly complex and sophisticated cognitive stages through which children pass from birth to adolescence (ontogeny) has been compared to the cognitive growth of humans as a species (phylogeny). Very simply, Jean PIAGET (1952, 1954) described cognitive development in children as a succession of stages whereby new schemes or ways of interacting with the environment are assimilated into the existing schemes in such a way as to produce qualitatively different cognitive capabilities (see Table 1). Very young infants have sensorimotor intelligence, according to Piaget, meaning that their knowledge about the world is confined to what they have learned through direct motoric interactions with objects and people around them, but they are incapable of abstract thought. At about 1½ years of age, toddlers become capable of preoperational thought, at which time they begin to

learn language and are able to organize their thoughts and their knowledge into various rules and classes. Not until about age 7 are children able to engage in concrete operational thought, reasoning deductively about phenomena that are concrete. Finally, in adolescence, the emergence of formal operational thought allows the use of abstract images and thought. This final stage is considered by Piaget to be the highest stage of cognitive development, a stage that only humans are capable of achieving. Each of Piaget's stages builds on, and supersedes, the previous stage, although cognitive processes from earlier stages may be evident during stressful or cognitively taxing situations. (Although Piaget's theory is not universally accepted in its entirety, most psychologists would accept at least some element of his point of view.)

This characterization of increasingly complex cognitive ability over time has led some researchers to hypothesize that the progression of cognitive abilities parallels the progression that occurred evolutionarily for our species (Gibson, 1990; Parker, 1985). S. T. Parker and K. R. Gibson (1979) have hypothesized that the intellectual and neocortical stages of development in today's human children recapitulate the stages of intellectual and neocortical growth throughout human

TABLE 1
Piaget's stages of cognitive development

<i>Cognitive Stage</i>	<i>Age</i>	<i>Description</i>
Sensorimotor period	Birth–18 mo.	Infant learns about world through motoric interactions; learns to differentiate self from the world; begins to understand cause and effect.
Preoperational period	18 mo.–7 yrs.	Child begins to use symbols and represent one thing with another; begins to use language; becomes better able to take another's perspective.
Concrete operational period	7 yrs.–12 yrs.	Child becomes able to use logic on observable and manipulable objects; becomes proficient at taking another's perspective.
Formal operational period	12 yrs. +	Adolescent can use abstract thought; can reason on abstract or verbal statements; can reflect on own thinking and still consider others' viewpoints.

evolution. Therefore, understanding how children learn and grow intellectually may help us understand how human intelligence evolved.

As support for their theory, Parker and Gibson (1979) stated that the common ancestor of great apes and humans displayed a rudimentary form of intelligence similar to that of children 1–4 years old. This intelligence arose as the first hominids used tools for extractive foraging (removal of embedded food, such as cracking open a nut) to obtain food. The descendants of the first hominids, in turn, displayed intelligence similar to that of children 4–7 years old. This intelligence was an adaptation to the complex tool-using involved in hunting, such as aimed-missile throwing, stone-tool manufacturing, and shelter construction.

Parker and Gibson's theory has received support, but others have pointed out a variety of situations in which the growth and development of the intellect of children today could not possibly follow the same pattern as the evolution of intelligence in human history. For example, C. J. Brainerd (1979) stated that the ability to understand and produce speech appears in human children long before the ability to throw objects accurately, while the course of evolution must have been exactly the opposite.

Thus, although it is interesting and possibly useful to consider the ontogeny of human intelligence as a recapitulation of its phylogeny, it is clear that this view is not without flaws. The forces that were acting on the human species during its evolution are not the same forces that are acting on children today during their intellectual growth from sensorimotor intelligence to formal operations. Children today generally do not hunt in the wild for their food, and the tools that they construct are usually not necessary for survival. Those forces that were at work during hominid evolution must be studied in order to understand the evolution of human intelligence. Several theories that incorporate such forces are reviewed below.

THEORIES OF BRAIN GROWTH AND REORGANIZATION

A number of factors have been proposed as the impetus for the evolution of the human brain (and consequently intelligence), including hunting (Ardrey,

1961) and food sharing (Isaac, 1979). Many of these theories share common foci by hypothesizing either tool use or social interaction as the prime mover of the adaptive evolutionary trend.

Proponents of the tool-use theory have suggested that tool using, toolmaking, and the related social requirements of our ancestors' hunting mode of life created powerful selection pressures for new intellectual abilities, such as planning, memory, and more complex and more efficient communication. Hominids with these abilities were more likely to survive and reproduce, passing on the abilities to the next generation (see Mayr, 1970; Tobias, 1967).

A number of theorists do not support the tool-use theory at all, instead viewing social interaction as the prime mover in the development of human intelligence. R. W. Byrne and A. Whiten (1988) have hypothesized that social behavior is more intellectually demanding than tool use and is thus more likely to be the prime mover. Many postulated scenarios incorporate different aspects of social interaction as the impetus for the evolution of human intelligence. For example, Byrne and Whiten (1988) have proposed a "Machiavellian expertise" explanation of brain evolution, wherein clever individuals relentlessly selected cleverness in their mates and more clever individuals had a greater survival advantage and were more likely to reproduce. This selection produced a powerful spiraling effect, with each generation exhibiting increasing levels of intelligence, until the human brain reached its present size.

Vigorous debate exists on the question of whether tool use or social behavior has played the more substantial role in the evolution of human intelligence. The myriad evidence and arguments for each theory (and various combinations of the two) are beyond the scope of this article. Perhaps most important is that many theorists have rejected the polarization of the social and technical (tool-use) theories, arguing that social structure, language, and advanced tool use form a single and unique adaptive complex. In humans, no single component could survive and advance without the others (Gibson, 1991). For example, our ancestors used language as a highly efficient form of expression to share new tool-using techniques among themselves and to pass them on to successive generations.

ADAPTATION VERSUS PREADAPTATION

Certain theorists have hypothesized that tool use and the social requirements of hunting are not sufficient to explain the evolution of human intelligence. K. R. Fialkowski (1986) cited an incongruity between the rate of acceleration in brain volume and the corresponding rate of sophistication in tool use. In the Lower and Middle Pleistocene, the brain doubled in size, yet no similar breakthrough was made in tool complexity. Fialkowski therefore theorized that the increase in the volume of the human brain was a "preadaptation," that is, a side effect of unknown adaptation that had nothing to do with intelligence. Preadaptation was the only way, Fialkowski maintained, to explain why the brain of human ancestors did not develop into a simpler one, sufficient for a pack-hunting predator, such as a wolf. The human brain, as it developed, far exceeded predator requirements.

Fialkowski suggested that the initiator of changes in the brain was the stress of heat involved in persistence hunting, the lengthy chasing of prey on the hot savannas. Brain tissue is sensitive to heat, and a rise of only 4° or 5° Celsius can create malfunctions in brain cells. The early human hunters had developed various cooling mechanisms for long-persistence hunting (e.g., loss of fur, body-cooling evaporation). However, these mechanisms were insufficient for long runs at higher temperatures. Brain volume and the number of neuronal connections therefore increased to improve the reliability of the brain, so that even if some cells malfunctioned, enough would remain to sustain vital functions. This theory required that the brain structures that changed the most should have been those that maintained vital functions, because those areas would need the increased number of neurons and interconnections to remain reliable when some neurons malfunctioned under stressful heat conditions. However, the exact opposite was the case; the change was primarily in association areas, not in those subserving vital functions.

D. Falk's (1990) radiator theory could perhaps resolve this incongruity. Falk suggested that bipedalism (walking on two feet) resulted in the development of a "radiator" for the brain that enabled the rapid increase in the brain size of human ancestors. Briefly, bipedalism caused a rearrangement of cranial blood

vessels. This new arrangement cooled the brain far more efficiently than blood-flow systems in the brains of other primates, enabling the brain to grow in size without becoming overheated. Falk also stated that thermal stress is more problematic to some parts of the human brain than to others. The cerebral cortex, which is on the outer surface of the brain, could possibly be more susceptible to damage from increased temperatures than are structures deeper in the brain. Cortical areas (including the prefrontal and association areas) would therefore grow more neurons and neuronal interconnections than would any other part of the brain because they were more at risk under stressful heat conditions and would need the growth to maintain reliability of their functions (Fialkowski, 1990).

A number of critics of preadaptation theories argue that such theories focus merely on the removal of brain-size constraints (like heat stress). A number of other constraints on brain size (such as metabolic-nutritional requirements and skull size) have also been pointed out (Barton, 1990). Critics charged that the removal of all constraints was necessary, but not sufficient, for the evolution of the human brain; it did not provide the driving force behind this evolution. This point brings us back to the original question (with a slight revision related to constraints): Once constraints on brain size were removed, what factors actively selected for the growth of the human brain and human intelligence: tool use, hunting, social interaction, a combination of these, or forces as yet unknown?

The array of proposed causal factors of the evolution of human intelligence are varied and interesting. Consensus has not been reached on exactly what facets of the organization of the brain have caused our unique type of intelligence, let alone the formative events that led to the brain changes. Several distinct factors, including foraging and social manipulation, are likely to be equally and perhaps interactively important, with no one cause sufficient to explain our intellectual evolution. These questions are likely to remain subjects of study for anthropologists and comparative psychologists for years to come.

CONCLUSION

Our intelligence as humans makes us unique as a species, different from any other in terms of our level of intelligence. Although the degree to which our in-

telligence has evolved may differ quantitatively from the degree of evolution among other animals, such as chimpanzees and dolphins, it nonetheless sets us apart qualitatively in terms of our quality of life and our mastery of our environment. Abstract reasoning, which we achieve when we are cognitively capable of formal operations at approximately age 12 (Piaget, 1954), allows us to make a cognitive leap from simply observing and acting on the world around us to actually imagining the world as it could be and hypothesizing and responding accordingly. Even before that, at approximately age 2, we begin to acquire a language system that is sufficiently complex to allow efficient communication among ourselves.

If, as Piaget has suggested, formal-operational thinking and abstract reasoning are the most advanced stages of intelligence in humans today, then it is easy to assume that we have reached the pinnacle of our intellectual evolution. D. A. Taylor (1982) took this concept to new heights by suggesting that our ability to think has allowed us to eliminate disease and hunger, allowing both the weak and the strong to survive and reproduce equally, thereby preventing genetic improvements across generations; he thus concluded that "evolution appears to be over." Unfortunately, disease and hunger have not been eliminated, so it is possible that even higher intellectual development will someday occur.

(See also: ANIMAL INTELLIGENCE: PRIMATE.)

BIBLIOGRAPHY

- ARDREY, R. (1961). *African genesis*. London: Collins.
- BARTON, R. A. (1990). Brain evolution in *Homo*: The "hood" theory. *Brain and Behavioral Sciences*, 13, 345–346.
- BRAINERD, C. J. (1979). Recapitulationism, Piaget and the evolution of intelligence: Déjà vu. *Behavioral and Brain Sciences*, 2, 381–382.
- BYRNE, R. W., & WHITEN, A. (1988). *Machiavellian intelligence: Social expertise and the evolution of intellect in monkeys, apes and humans*. New York: Clarendon Press/Oxford Science Publications.
- FALK, D. (1990). Brain evolution in *Homo*: The "radiator" theory. *Behavioral and Brain Sciences*, 13, 333–338.
- FALK, D. (1992). *Braindance*. New York: Holt.
- FIALKOWSKI, K. R. (1986). A mechanism for the origin of the human brain: A hypothesis. *Current Anthropology*, 27, 288–290.
- FIALKOWSKI, K. R. (1990). On the origin of the brain and heat stress: New facts. *Current Anthropology*, 31, 187–188.
- GIBSON, K. (1991). Tools, language and intelligence: Evolutionary implications. *Man*, 26, 255–264.
- GIBSON, K. (1990). New perspectives on instincts and intelligence: Brain size and the emergence of hierarchical mental constructionist skills. In S. T. Parker & K. R. Gibson (Eds.), *Language and intelligence in monkeys and apes*. Cambridge: Cambridge University Press.
- HOLLOWAY, R. L. (1966). Cranial capacity, neural reorganization, and hominid evolution: A search for more suitable parameters. *American Anthropologist*, 68, 103–121.
- HOLLOWAY, R. L. (1968). The evolution of the primate brain: Some aspects of quantitative relationships. *Brain Research*, 7, 121–172.
- ISAAC, G. L. (1979). Evolutionary hypotheses. *Behavioral and Brain Sciences*, 2, 388.
- JERISON, H. J. (1973). *Evolution of the brain and intelligence*. New York: Academic Press.
- JERISON, H. J. (1982). The evolution of biological intelligence. In R. J. Sternberg (Ed.), *Handbook of human learning*. Cambridge: Cambridge University Press.
- MAYR, E. (1970). *Population, species and evolution*. Cambridge, MA: Harvard University Press.
- PARKER, S. T. (1985). Higher intelligence as adaptation for social and technological strategies in early *Homo sapiens*. In G. Butterworth, J. Rutkowska, & M. Scaife (Eds.), *Evolution and developmental psychology*. New York: St. Martin's Press.
- PARKER, S. T., & GIBSON, K. R. (1979). A developmental model for the evolution of language and intelligence in early hominids. *Behavioral and Brain Sciences*, 2, 367–408.
- PIAGET, J. (1952). *The origins of intelligence in children*. New York: Norton.
- PIAGET, J. (1954). *The construction of reality in the child*. New York: Basic Books.
- PLOMIN, R.; DEFRIES, J. C.; & MCCLEARN, G. E. (1990). *Behavioral genetics: A primer* (2nd ed.). New York: W. H. Freeman.
- STERNBERG, R. J., & SALTER, W. (1982). Conceptions of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. Cambridge: Cambridge University Press.
- TAYLOR, D. A. (1982). *Mind*. New York: Simon & Schuster.

TOBIAS, P. V. (1967). Cultural hominization among the earliest African Plio-Pleistocene hominids. *Proceedings of the Prehistoric Society*, 33, 367–376.

LISABETH F. DiLALLA
CAROL PATRICK

EXPERTISE See KNOWLEDGE.

EXPERIMENTAL DESIGN All research begins with questions: How many kinds of intelligence are there? How much do people vary in intelligence? How much variation is due to genetic factors, or to specific environmental factors (such as prenatal drug use)? Can intelligence be changed? The purpose of research design is to plan a study that will enable the question of interest to be answered as fully and exactly as possible, with the least equivocation and the fewest possible unverified assumptions. Although some research designs are simple, most are not because it is usually necessary to eliminate various plausible rival hypotheses.

This article describes several types of research designs: experimental, quasi-experimental, correlational, longitudinal and repeated measures, and twin and family studies. Each has advantages and disadvantages, and some are useful for answering specific kinds of questions.

Experimental designs (and attempts to approximate them, called *quasi-experimental designs*) attempt to make fair comparisons of groups of people that were treated differently; these designs therefore emphasize the elimination of plausible rival hypotheses about why the groups differ on some characteristic of interest. Other types of studies (e.g., surveys) are more descriptive, and for these studies a primary goal is to obtain a sample that is representative of the population one wishes to describe. In the terminology of Campbell and Stanley (1966), the experimental study has internal validity as its primary criterion; the interest is in whether the treatment had an effect on the people in the study. The descriptive study has external validity as its primary criterion: the interest is in the applicability of the results to a specific population.

The research design process includes construction of measurement procedures and instruments, selection

of subjects to be studied, assigning subjects (in a randomized study) to different treatments (or measurement conditions), collection of data, and statistical analysis. Statistical analysis goes hand-in-hand with research design; one should never design a study without knowing how to analyze the data. While statistical issues will be discussed only minimally here, anyone who designs research should realize that for every research question, there must be a statistical hypothesis tested or an unknown quantity (*parameter*) estimated. If one cannot specify the statistical methods that would be used to answer the research question, then the research question may be specified too vaguely and thus may be unanswerable.

MAJOR CATEGORIES OF STUDIES

Randomized Experiments. The research design that provides the most unequivocal evidence is the true experiment, in which individuals are randomly assigned to different treatments. If we are interested in whether cognitive training can increase intelligence, then we could randomly assign some people to receive training and others not to receive it. By assigning people at random, we make it extremely unlikely that the two groups differ systematically from each other before the training period begins. Therefore, if the groups differ in intelligence after the first group (typically called the *experimental*, or *treatment*, group) has been trained, and the other (called the *control* group) has not, we can be reasonably sure that the difference is because of the training, and not because the groups differed before training. While only the simplest design was illustrated here, a later section describes designs that deal with some complications that can arise in the research process.

Quasi-Experimental Designs. Quasi-experimental designs, like experimental designs, involve comparisons of groups of people that received different treatments, but do not use randomization to determine who is in each group. For example, suppose we did not randomly assign some people to receive experimental treatment, but just measured some who happened to choose to take a course that promised increased cognitive abilities, and we measured some others who did not take such a course. A comparison of these groups would prove little, because we would

not know whether a difference was a result of training, or of the tendency of people with higher (or lower) intelligence to take such courses. Therefore, preexisting differences between the groups would be a *plausible alternative hypothesis* to the hypothesis we want to test, which is that the cognitive training has an effect. One goal in designing research is to eliminate as many plausible alternative explanations of the results as possible, so that the only plausible remaining explanation is the one the researcher wants to test.

Although randomization is usually considered the best way to eliminate alternative explanations for group differences, nonrandomized studies are not necessarily worthless. Some quasi-experimental designs have more controls for plausible alternative hypotheses than do other designs. For example, suppose that subjects were not randomly assigned to groups, but that each subject was given an intelligence test before the training period (a *pretest*) as well as afterward. This simple addition to the design eliminates or reduces the likelihood of many otherwise plausible explanations. For example, if the groups were not equal at the start, the extent of the initial difference can now be measured and accounted for.

Other general strategies for eliminating plausible alternative hypotheses are matching, statistical control, and selection modeling. *Matching* is done by finding, for each person getting treatment, another person who is similar to the first person, but who did not get the treatment. By the use of this technique, one can reduce the bias in comparing the two groups. Unfortunately, it is not possible to match pairs of people on *all* important factors, because it is impossible to know all important factors, let alone measure them perfectly. Omission of important factors for matching, and errors of measurement, will both contribute to bias in the estimate of a treatment effect.

Another approach is to measure as many relevant influences on the outcome as possible. The research then constructs a statistical model to predict the outcome from knowledge of treatment status and these other potential influences (called *control* or *contrast variables*). The statistical model provides an estimate of the effect of treatment for people who are the same on all control variables.

If the outcome of interest is cognitive ability in some area, then any prior measure related to that abil-

ity would be a useful control variable. In the best case, one would have a pretest score; in a worse case, perhaps only the number of years of education. This method is analogous to matching, but is much simpler to accomplish than actually having to search through the pool of control subjects to find a match for each experimental subject. As with matching, plausible alternative hypotheses usually involve omitted control variables and imperfect measurement of control variables that are included.

A third strategy, commonly called *selection modeling*, is to construct a statistical model to predict who will get the treatment and who will not. This may be possible even though the researcher has no control over who gets the treatment. If all important factors are included in such a model, then the selection process can be accounted for statistically in calculating an unbiased estimate of the effect of treatment. As with the other strategies one can seldom be sure that all important factors have been included.

Correlational Designs and Causal Models.

Suppose that one is interested in determining how strongly intellectual development is affected by various environmental variables. Many such factors cannot easily be manipulated: parents' educational levels, occupation, family income, characteristics of the neighborhood, and quality of schooling. Another complication is that some of these variables may affect each other as well as affecting intellectual development. Even worse, it is often possible to measure important influences not when they occur, but only after the fact (*ex post facto*). Yet in spite of these problems, a statistical method called *path analysis* can be used to test various theories about which variables are important influences on each other and on intellectual development.

A common dictum in the social sciences is that "correlation does not imply causation," which means that two variables can be related to each other (correlated) without any necessary causal connection between them. While this is true, it is also true that causation implies correlation; if one variable has a causal influence on another, then the two will be correlated (all other things being equal). It is this rule that is used to advantage by path analysis: Different models are specified about which variables affect which other variables. By examining the actual pattern of correla-

tions among the variables, the researcher can eliminate models whose predictions are not consistent with what is observed. Models that are consistent with what is observed are more plausible than those that are inconsistent with the data, although it is never possible to prove that a particular model is correct.

Longitudinal Studies. Many research questions involve the course of growth and development. In young children, for example, one might measure vocabulary size as a crude index of intellectual development. If groups of children of different ages were measured, we could see what the average child was like at each age, but if we want to see how individual children grow, then we must measure the same children at different ages. A longitudinal study is one in which individuals are measured at several ages, and usually involves the same or similar measurements at each age. Longitudinal studies allow the examination of individual growth, and of how growth patterns differ with stable characteristics—for example, how vocabulary growth in children varies with the educational level of the mother.

Comparative Studies. Do people from different countries, cultures, or subcultures process information in different ways? To what degree is the development of cognitive abilities affected by changes in a culture? Comparative studies are designed to answer such questions. Methodological difficulties abound for researchers doing comparative studies, because so many plausible alternative hypotheses about the reasons for differences exist, including language differences.

Twin and Family Studies. To determine whether a trait is influenced by genetic factors, and if so, to what degree, people who differ in degrees of relatedness must be studied to see whether the degree of similarity in the trait differs with degree of genetic relatedness. The most usual such study involves identical twins; because they have identical genetic material, differences between them must be due to environmental influences (see TWIN STUDIES).

COMMON VARIATIONS IN DESIGNS

Factorial Designs. Researchers are often able to answer more than one research question in one study. To expand on the original example of the eval-

uation of a cognitive training program, suppose that the researcher also believes that a new drug has some prospect of increasing cognitive performance. One approach might be to do two experiments; one would evaluate the cognitive training program, and the other would evaluate the utility of the drug. This approach, however, would not only double the number of subjects needed, but would also lose the ability to answer questions such as whether the cognitive training is equally effective for those who take the drug and those who do not. Questions such as these are answered statistically by the search for *interactions*, as will be explained below.

A study that would answer both questions at the same time, as well as the question of whether there is an interaction, would have participants assigned at random to one of four groups. One group would get cognitive training and the drug; one would get only cognitive training; one would get only the drug; and one would get neither. This type of design is known as a *factorial design*; the two *factors* in this study are cognitive training and drug. The *levels* of the factors are “yes” and “no,” since for each factor, people either do or do not get the treatment. Because each of the two factors has two levels, the design is known as a 2×2 design; the notation reminds the researcher that there are 2 times 2, or 4, conditions (also called *cells*) in the study.

The outcome variable in this study would be a score on some type of mental test. For each of the four groups, we would calculate the mean (average) score; statistical tests would determine whether or not differences among these means were most likely due to the treatment. As an example of the possible results, the means of the four groups might be as listed in the following diagram:

	Drug	
Cognitive Training	Yes	No
Yes	70	60
No	50	40

Looking first at the question of whether cognitive training had an effect, one can average the groups who did and did not get the drug, and find that those who had cognitive training obtained an average of $(70 +$

$60/2 = 65$ on the test, while those who did not have cognitive training obtained an average of $(50 + 40)/2 = 45$. Therefore, cognitive training appears to have raised scores by an average of 20 points. Next, looking at the effect of the drug, one sees that those who were given the drug averaged $(70 + 50)/2 = 60$, while those who were not given the drug averaged $(60 + 40)/2 = 50$. The drug apparently increases performance by 10 points.

Finally, we examine whether the effect of the drug is the same for those who did get cognitive training as for those who did not. For those who did, the drug increased performance by $70 - 60$, or 10 points, while for those who did not, the drug increased performance by $50 - 40 = 10$ points. Because the effect of the drug was the same regardless of whether or not cognitive training was received, there is no *interaction* between the effect of drug and the effect of training.

Another way of saying this is that the effects of drugs and cognitive training are *additive*; if you get cognitive training *and* the drug, your score increases by 20 points for cognitive training and by 10 points for getting the drug, or by 30 points in all. A typical person who gets neither cognitive training nor the drug is expected to get a score of 40, while someone who gets both is expected to get a score 30 points higher, or 70. (To see what would happen if there were an interaction, change any one number in the table; for example, change the mean for those getting both cognitive training and the drug to 60.)

Sampling Issues. Among the basic decisions to be made in any research are how many subjects to use, and how to allocate them to different treatment groups. In surveys and comparative studies, it is also important to consider how subjects are to be selected from the population to which generalization is desired. In general, the more subjects are included in a study, the more precisely we can estimate important quantities such as means, and the more easily we can detect differences among groups. Too few subjects leads to imprecise estimates and an inability to tell whether groups differ. Too many subjects wastes resources, enabling us to be more precise than we need to be, and to detect differences that are too small to be of practical importance. The ability to detect a difference between groups is measured by a quantity called *statistical power*.

Although researchers can always increase power by including more subjects in a study, less costly methods are often available. If we could match together pairs of subjects who, before training, are similar in intelligence, and randomly assign one person in each pair to get the experimental treatment and the other to be the control (receiving no treatment), then the power of the study to detect effects will be increased. This is called a *randomized blocks* design, because small blocks of subjects are formed who are similar on a variable related to the outcome, and members of each block are then randomly assigned to various treatment conditions.

Stratification on potentially important predictors of the outcome variable is another tool for increasing power, but in addition, it allows testing for interactions in a manner similar to factorial experiments. For example, suppose we classify students by the level of education of their mothers (or fathers), using categories such as “less than 4 years of high school,” “high school graduate,” “college graduate,” and “professional degree.” One can see not only whether education is related to the outcome of the study, but also whether the treatment has the same effect for people whose parents had different amounts of education; if not, there is an interaction between education and treatment condition. Even if there is no interaction, stratification is advantageous because the power to detect effects is increased when the stratifying variable is related to the outcome.

Incomplete Designs. Special designs have been developed to handle particular problems. For example, an experimenter may wish to include a large number of factors in a design. If there are 10 factors, each with only 2 levels, there would be 1,024 cells (conditions) in the design, requiring at least 2,048 subjects if there were only 2 subjects per cell. To reduce the number of subjects to a manageable number, experimenters can choose to use only some of the 1,024 cells of the design, and in extreme conditions to use only one subject per cell. Such a design, called a *fractional factorial* design, makes an assumption that the most complicated interactions that might exist in the design in fact are of negligible size.

As another example, one might have many schools that can be included in a study, but with only a few students might be available in each school. In fact,

there may be fewer students per school than there are cells in the design. An *incomplete blocks* design would be useful in this situation, enabling the researcher to be sure that treatment effects can be separated from the differences among schools.

Traditional experimental design considers only the design of the current experiment; design in industrial production has expanded to consider how to design sequences of experiments to optimize the production process, with the design of each experiment depending on the results of the previous experiment. Such sequences of designs might be of use in discovering how to optimize intelligence with the least cost of experimentation.

FURTHER READING

Classic sources on important considerations in experimental and quasi-experimental design include Campbell and Stanley (1966), Cochran (1983), and Cochran and Cox (1957). The last is like many contemporary sources in that statistical as well as design considerations are an integral part of the text. Most recent texts on experimental design and analysis differ from prior works—not in the types of designs discussed but in the use of a more comprehensive statistical approach via general linear models. Typical of the newer approach are books by Woodward, Bonnett, and Brecht (1990) and Maxwell and Delaney (1990). An excellent nontechnical work illustrating research design and the interpretation of statistical tests is Huck, Cormier, and Bounds (1974). The design of sequences of experiments to optimize industrial output is illustrated in the work of Box and Draper (1969). Maddala (1983) discusses selection models, as well as other useful new approaches to design and analysis developed in the field of economics.

BIBLIOGRAPHY

- BOX, G. E. P., & DRAPER, N. R. (1969). *Evolutionary operation: A statistical method for process improvement*. New York: Wiley.
- CAMPBELL, D. T., & STANLEY, J. C. (1966). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- COCHRAN, W. G. (1983). *Planning and analysis of observational studies*. New York: Wiley.
- COCHRAN, W. G., & COX, G. M. (1957). *Experimental designs* (2nd ed.). New York: Wiley.
- HUCK, S. W., CORMIER, W. H., & BOUNDS, W. G., JR. (1974). *Reading statistics and research*. New York: Harper & Row.
- MADDALA, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge: Cambridge University Press.
- MAXWELL, S. E., & DELANEY, H. D. (1990). *Designing experiments and analyzing data*. Belmont, CA: Wadsworth.
- WOODWARD, J. A., BONETT, D. G., & BRECHT, M. (1990). *Introduction to linear models and experimental design*. San Diego, CA: Harcourt Brace Jovanovich.

DAVID RINDSKOPF

EYSENCK, HANS J. (1916–) One of the most famous and controversial psychologists of the latter half of the twentieth century and the leading exponent of the London school of biological and quantitative psychology established by Francis GALTON, Charles SPEARMAN, and Cyril BURT, Hans Jurgen Eysenck has made prolific and influential empirical and theoretical contributions to differential psychology, most notably in the area of research on personality and human mental ability. In his studies of personality, he applied the quantitative methods developed by the London school, particularly factor analysis, along with the hypothetico-deductive use of experimental methods involving the constructs of Ivan Pavlov and Clark Hull. Eysenck's prolific research is cited in many areas of the psychological literature, including extraversion, neurosis, behavior therapy, critiques of psychoanalysis and Freudian theory, sexual behavior, the psychology of politics, smoking and health, measurement and theory of intelligence, behavioral genetics, race differences, and creativity and genius. He has even examined parapsychology and astrology from an objective and scientific standpoint. Modgil and Modgil (1986) have edited a fairly comprehensive volume of critical commentaries on Eysenck's varied contributions.

Born in Berlin, Eysenck was the only child of comfortably well-off, cultured parents. His father was a stage actor, his mother a movie actress. Graduating from the Gymnasium in Berlin in 1933, the year that Hitler came to power, he left Germany in protest and

spent a summer in England at the University of Exeter, followed by a few months in France at the University of Dijon. In 1934 he enrolled at University College, London, majoring in psychology under Sir Cyril Burt. He received his bachelor of arts degree in 1938 and his doctorate in 1942, whereupon he was appointed chief psychologist at the Mill Hill Emergency Hospital (London) during World War II (1942–1945). He then became psychologist at the Maudsley Hospital, London's leading psychiatric facility (1945–1950), followed by promotion to reader in psychology and director of the Psychology Department of the Institute of Psychiatry of the University of London (1950–1955). From 1955 to 1984 he was professor of psychology at the institute. As emeritus professor since 1984, he remained as active as ever, researching, writing books and articles, giving invited lectures around the world, and editing the international journal that he founded in 1980, *Personality and Individual Differences*.

Eysenck's contribution to intelligence consists of his own considerable research output in addition to the strong theoretical influence he has had on his colleagues and on many other researchers who have made significant contributions. Eysenck takes a "hard science" approach, viewing intelligence not as a thing or a denotative noun, but as a theoretical construct similar to the basic concepts of physics, for example, mass, gravitation, and potential energy. He insists that neither the subjective nor the behavioral manifestations of intelligence—reasoning, memory, learning, problem solving, and the like—can constitute a proper definition of intelligence nor does Spearman's g (general ability), which merely reflects the fact of individual differences in intelligence. Rather than being a definition or an explanation, g is a phenomenon itself in need of explanation. This must involve constructs beyond subjective and behavioral phenomena. While acknowledging the importance of factor analysis for analyzing the correlational structure of abilities represented in a battery of diverse tests and for measuring independent components of mental ability, such as g and various group factors, Eysenck was among the first to recognize the impotence of factor analysis for understanding the causal basis of intelligence differences. The causal question, he argued, must appeal to the methods of behavioral genetics and neurophysiology.

Following D. O. HEBB, Eysenck emphasizes the essential distinction between three classes of phenomena associated with cognitive performance, referred to as Intelligences A, B, and C. Intelligence A is the biological substrate of mental ability, the brain's neuroanatomy and physiology. Intelligence B is the manifestation of Intelligence A and everything that influences its expression in "real life" behavior. Intelligence C (first so labeled by P. E. Vernon) is the level of performance on psychometric tests of cognitive ability. Eysenck dismisses Intelligence B as unsuitable for scientific study because it represents such a complex interaction of Intelligence A with variation in a host of cultural, educational, and other social and psychological influences in the course of the individuals' development, as well as being confounded by personality and motivation, thereby making it (Intelligence B) essentially unmeasurable and unamenable to the purposes of scientific formulation. Intelligence C, however, being based on psychometric tests of ability, does allow quantitative and statistical treatment of data and is indeed essential for the study of mental ability. Tests vary widely in the degree to which they reflect Intelligence A or Intelligence B, however. Verbal tests with culturally and educationally loaded items, for instance, are closer to Intelligence B, while certain nonverbal tests of reasoning and problem solving using simple pictures or geometric forms that are highly familiar to all examinees may better reflect Intelligence A. The components of variance in reaction time and inspection time that are correlated with Intelligence C probably come even closer to Intelligence A, and physiological measurements derived from the average evoked potential, the rate of glucose uptake in the brain detected by positron emission tomography (PET), and nerve conduction velocity in the brain (which are all correlated with intelligence quotient [IQ]) are the closest to Intelligence A. In the latter part of his career, Eysenck focused on the empirical relation between Intelligence C and its biological basis, or Intelligence A.

In the 1950s, Eysenck revived Galton's hypothesis that mental speed is what underlies individual differences in g . Measurement of the time that individuals take to solve single test items of varying difficulty permitted the analysis of test performance into three main sources of variance: speed, continuance (i.e., persistence of effort), and error checking. Because only

speed can be truly regarded as a cognitive variable, the other two variables really being aspects of personality, Eysenck and his coworkers focused their research on mental speed, as measured by choice reaction time, inspection time (a measure of purely perceptual speed), and the latency and waveform of brain-evoked potentials. As evidence accumulated showing that trial-to-trial intraindividual variability in these measures is more highly correlated (negatively) with g than is speed itself, Eysenck promulgated the hypothesis that g reflects the rate of errors in the neural transmission of information through the cortex. In other words, the level of a person's intelligence depends on the probability that neurally encoded messages will be transmitted to their destinations in the brain without degradation or distortion by random "noise" in the nervous system. This theory has some empirical support from studies of reaction times and evoked potentials and is under continuing investigation.

(See also: EEG EVOKED POTENTIALS.)

BIBLIOGRAPHY

- CARLSON, J. S., & WIDAMAN, K. F. (1986). Eysenck on intelligence: A critical perspective. In S. Modgil & C. Modgil (Eds.), *Hans Eysenck: Consensus and controversy*. Philadelphia: Falmer.
- EYSENCK, H. J. (1967). Intelligence assessment: A theoretical and experimental approach. *British Journal of Educational Psychology*, 37, 81–98.
- EYSENCK, H. J. (1979). *The structure and measurement of intelligence*. Heidelberg: Springer-Verlag.
- EYSENCK, H. J. (ED.). (1982). *A model for intelligence*. Heidelberg: Springer-Verlag.
- EYSENCK, H. J. (1986). The theory of intelligence and the psychophysiology of cognition. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3, pp. 1–34). Hillsdale, NJ: Erlbaum.
- EYSENCK, H. J. (1991). *Rebel with a cause*. London: Routledge. Autobiography.
- GIBSON, H. B. (1981). *Hans Eysenck: The man and his work*. London: Peter Owen.
- JENSEN, A. R. (1986). The theory of intelligence. In S. Modgil & C. Modgil (Eds.), *Hans Eysenck: Consensus and controversy*. Philadelphia: Falmer.
- MODGIL, S., & MODGIL, C. (EDS.). (1986). *Hans Eysenck: Consensus and controversy*. Philadelphia: Falmer.

ARTHUR R. JENSEN

F

FACET THEORY Formulated by Louis GUTTMAN (1959, 1991), facet theory (FT) is fundamental to the study of human intelligence. The central notion of facet theory (FT), the *facet*, classifies the elements of a domain of interest into types. The facet “gender,” for example, classifies persons into “males” and “females.” Similarly, the facet “behavior modality” classifies attitudinal behavior into “emotional,” “cognitive,” and “actional” behavior. The use of several facets at the same time partitions a domain of interest into multifaceted types. Intelligence items in FT, for example, are defined as questions that ask about an individual’s behavior and assess it on a scale from “very right” to “very wrong” according to an objective rule (Guttman, 1965). A particular case of such items consists of the tests in paper-and-pencil intelligence test batteries. Among other challenges, these tests require the test taker to find verbal analogies, solve arithmetic problems, and identify patterns that complete series of figures. Hence, they can be classified by the facet “language of presentation” into “numerical,” “verbal,” and “geometrical” ones. Because these tests relate to different abilities, a second facet, “required mental operation” arises. It classifies tests into those in which the test taker has to “infer,” “apply,” or “learn” a rule (Guttman & Levy, 1991). In combination, the two facets “language of presentation” and “required mental operation” partition intelligence items into nine types,

such as numerical tests requiring the inference and application of a rule and geometrical tests requiring the learning of a rule.

In FT, facets are typically not just listed but rather expressed in the framework of a mapping sentence, which shows the roles the facets play relative to each other and relative to what is being observed (the range). Consider the following statement: “Person (p) performs on a task presented in (verbal, numerical, geometrical) language and requiring (learning, application, inference) of an objective rule \rightarrow (very right, very wrong) according to that rule.” The terms in parentheses are the facets. The persons are not facetized further in this example. The questions are structured by the two facets from above, “requirement” and “language.” The range of the mapping sentence is the scale on the right-hand side of the arrow. The arrow symbolizes an observational mapping of every person crossed with every (doubly coded) test into the range (data).

The development of a mapping sentence typically starts with the range. For example, in assessing behavior in terms of “right to wrong according to an objective rule,” which defines intelligence behavior, the range shows what the items have in common. Facets such as “language of presentation” or “required mental operation,” on the other hand, denote how the items differ among each other. Well-designed research typ-

ically distinguishes thousands of item types because the number of item types increases greatly with the number of facets.

If the facets are clear and their roles in the range are well understood, mapping sentences provide conceptual clarity and control over a field of interest. The challenging hypothesis concerns the empirical control that they possibly provide. A traditional formulation of this hypothesis assumes that the facets of the questions should “explain” or “predict” the data in some way. The most direct prediction is to check whether the distinctions made by the facets are mirrored facet by facet by corresponding differences in the data. For example, tests that require the test taker to infer, apply, or learn a rule should lead to different responses of the person tested. One particular assumption of the “difference” is that in general inferential tests are most difficult and learning tests are least difficult. Application tests are in the middle.

A lesser hypothesis is that different item types fall into different regions of a geometrical representation of the data. As an illustration, Table 1 shows correlations of intelligence tests and *structuples*, that is, the codings of the tests regarding the facets “language” and “requirement” discussed above. These correlations are represented as a geometrical picture that is much more accessible than a direct presentation of correlation coefficients. One of the techniques that yields an easily understandable representation is multidimensional scaling (MDS), which first assigns to each test a particular point in space and then moves the points around until the distance between any two tends to

decrease as the correlation between the corresponding tests increases (Borg & Lingoes, 1987). Figure 1 shows an MDS configuration whose distances correspond very well to the correlations in Table 1. Notice that tests 1 and 2 have a correlation of .67 and are close together in the MDS space. Tests 1 and 5 have a correlation of .12 and are quite far apart. Hence, Figure 1 shows in geometrical terms what correlations in Table 1 reveal in numerical language. Figure 2 demonstrates that the MDS configuration can be divided such that each partitioning line splits it into two regions containing only points of one type. (Points of the N-type lie above the solid line, and points of the G-type below that line. The dotted line separates I-type points from A-type points.) The peculiar curvature of the dotted line was chosen because if all nine test types distinguished by the two facets were observed, the MDS configuration should form a pattern similar to a dart-board (see RADEX THEORY).

With many points or differentiated facets or both, a simple correspondence between regions and *structuples* is a remarkable finding. As mathematical demonstrations indicate, arbitrary assignments of *structuples* to the points do not lead in general to such a lawfulness. Partitionings with relatively smooth cutting lines are generally also more reliable. Moreover, they help clarify the roles the various facets play with respect to the data. Within MDS, such roles are often reflected in particular ways of cutting the space. In a plane, a polar facet, for example, generates a circular pattern of wedgelike regions. A modulating facet, on the other hand, leads to concentric bands. An axial

TABLE 1
Intercorrelations of eight intelligence tests, together with content *structuples*

<i>Structuple</i>		1	2	3	4	5	6	7	8
NA	1	1.00	.67	.40	.19	.12	.25	.26	.39
NA	2	.67	1.00	.50	.26	.20	.28	.26	.38
NI	3	.40	.50	1.00	.52	.39	.31	.18	.24
GI	4	.19	.26	.52	1.00	.55	.49	.25	.22
GI	5	.12	.20	.39	.55	1.00	.46	.29	.14
GA	6	.25	.28	.31	.49	.46	1.00	.42	.38
GA	7	.26	.26	.18	.25	.29	.42	1.00	.40
GA	8	.39	.38	.24	.22	.14	.38	.40	1.00

SOURCE: Guttman, 1965

N = numerical, G = geometrical; A = application, I = inference.

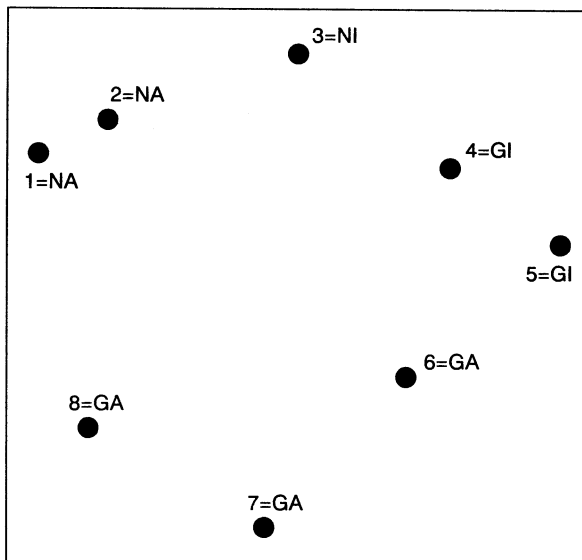


Figure 1
MDS configuration of correlations in Table 1. For meaning of structuples, see Table 1.

facet partitions the space into parallel stripes. If particular facets are reliably related to particular regional patterns, one has established a regional law.

Any replicable regionalization is interesting, but certain patterns have arisen so often that they have been given particular names (such as *simplex*, *circumplex*, *radex*, and *conex*). Because facet theory is so well established, new facets come to mind, allowing the formal derivation of the particular regional pattern from the properties of these facets. The qualitative facet “requirement” = (inference, application, learning), for example, does not explain why the corresponding regions emerge in an ordered way. Attempts were therefore made to replace “requirement” by the ordered facet “degree of complexity” (Guttman, 1954; Snow, Kyllonen, & Marshalek, 1984; Tziner & Rimmer, 1984).

MDS and regional examinations of the data are not the only data-analytic techniques in FT. Nonetheless, they show what all FT methods have in common. They are all nonlinear and “soft,” and they all evaluate differences in the data related to distinctions made in the definitional framework of the items. In contrast to more traditional methods, they do not assume a rigid scaffolding of items such as linearity, additivity, and normality. This difference typically allows a view of more general and more robust structures in the data.

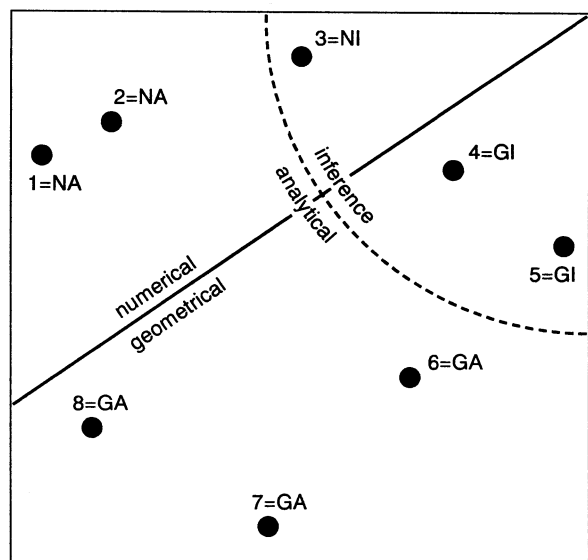


Figure 2
MDS configuration of data in Table 1. Lines partition space into G- vs. N-regions and into A- vs. I-regions, respectively.

Other facet theories exist, in particular J. P. GUILFORD’s intelligence theory, which Guttman regarded as a partial “application” of his own scheme. What is left open is a sharp definition of the universe of items—their range, in particular—and intrinsic correspondence hypotheses of item types to data. Guttman developed his own application of FT to intelligence theory, which goes beyond classifying intellectual tasks. It defines especially the range of intelligence items and states that such items should be correlated positively among each other (under certain side constraints). An illustrative case for this “first law of intelligence” is shown in Table 1. Another empirical law established in Guttman’s facet approach to intelligence is a regional law—the cylindrex structure of intelligence items (Guttman & Levy, 1991).

(See also: STRUCTURE OF INTELLIGENCE.)

BIBLIOGRAPHY

- BORG, I., & LINGOES, J. C. (1987). *Multidimensional similarity structure analysis*. New York: Springer-Verlag.
- GUILFORD, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- GUTTMAN, L. (1954). A new approach to factor analysis:

- The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258–348). New York: Free Press.
- GUTTMAN, L. (1959). Introduction to facet design and analysis. *Proceedings of the 15th International Congress of Psychology in Brussels* (pp. 130–132). Amsterdam: North-Holland.
- GUTTMAN, L. (1965). The structure of interrelations among intelligence tests. In C. W. Harris (Ed.), *Proceedings of the 1964 Invitational Conference on Testing Problems* (pp. 25–36). Princeton, NJ: Educational Testing Service.
- GUTTMAN, L. (1991). *Louis Guttman: In Memoriam—Chapters from an unfinished textbook on facet theory*. Jerusalem: Israel Academy of Sciences and Humanities.
- GUTTMAN, L., & LEVY, S. (1991). Two structural laws for intelligence tests. *Intelligence*, 15, 79–103.
- SHYE, S. (1985). *Multiple scaling*. Amsterdam: North-Holland.
- SNOW, R. E., KYLLONEN, P. C., & MARSHALEK, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3, pp. 47–103). Hillsdale, NJ: Erlbaum.
- TZINER, A., & RIMMER, A. (1984). Examination of an extension of Guttman's model of ability tests. *Applied Psychological Measurement*, 8, 59–69.

INGWER BORG

FACTOR ANALYSIS *Factor analysis* is a general term for: (1) a set of theoretical concepts used by scientists to describe hypothetical constructs; (2) a set of mathematical models used to express observations resulting from the action of unobserved or latent variables termed common factors; (3) a set of statistical techniques used to examine the goodness of fit of these theoretical models to the measured observations; (4) a set of computational techniques, largely developed by psychologists in research on individual differences in cognition and personality, and now widely used in a number of scientific disciplines.

The key idea in any factor analysis is that a large number of observed phenomena may be the result of a smaller number of unobserved or latent phenomena. This idea directs empirical studies to isolate a set of underlying common factors that represent parsimonious and reliable sources of differences between individuals and groups. This is one basic idea in common factors theories of intellectual abilities. It is similar to

the main areas of research and experimentation in many other areas of science—the “quarks and atoms” of physics, the “molecules and elements” of chemistry, the “genes and viruses” of biology, and the “unobserved planets and black holes” of astronomy. The purpose of this entry is to give a nontechnical overview of basic concepts in factor analysis.

INTELLIGENCE THEORY AND FACTOR ANALYSIS

Factor analysis techniques relate closely to some of the main theoretical issues in intellectual abilities. The factor analysis model has played an important role in determining how we view intellectual abilities (Sternberg, 1985). This history begins with the early statistical work of Francis GALTON on individual differences, of Yule and Pearson on statistical methods, and Charles SPEARMAN on mathematical and statistical models (Boring, 1950). Factor analysis of abilities continues to be an active research area today (Carroll, 1993).

The first factor analysis experiments were presented by Spearman (see TWO-FACTOR THEORY), who suggested that the construct of intelligence could be defined as a unitary function that was manifest in many different kinds of test responses. A *common factor*, labeled g , was defined as an unobserved or latent variable that has an influence on a broad set of observed variables. Common factors were further distinguished from unobserved unique factors (also called *specific factors*), which influence only one measured variable and are themselves uncorrelated. Spearman examined this idea scientifically by inventing the basic methods for testing the proportionality of response functions. This idea of a single intellectual function was a popular idea at the turn of the century but, until Spearman's work, it lacked a formal scientific expression (Gould, 1981; Carroll, 1993). Contemporary factor analysis techniques are still based on this separation of an observed variable into two kinds of unobserved factors.

Spearman's theory of a single g has been challenged by data from many factor analysis experiments. Cyril BURT (1909) proposed the use of additional “group factors”—these are defined as common factors like g but only for some subsets of the variables. Burt also

showed how group factors improved the goodness-of-fit to real data. L. L. THURSTONE (1931, 1947) pioneered the ideas and methods for the general extension of the group factors termed *multiple factor analysis*. Thurstone showed how a broad set of “primary mental ability” factors could well represent relationships in data. Other investigators using factor analysis presented evidence for HIERARCHICAL THEORIES OF INTELLIGENCE. A general two-factor alternative was proposed by R. B. Cattell (1941) as the “*Gf-Gc* theory” for FLUID AND CRYSTALLIZED INTELLIGENCE (Horn & Cattell, 1966), and he later expanded this into a “triadic theory of ability factors” (see TRIADIC THEORY OF ABILITY STRUCTURE). Carroll (1993) has presented a complete overview and reanalysis of much cognitive abilities data using factor analysis methods. Other summaries on intellectual abilities using contemporary techniques have been presented by Horn (1988), Gustafsson (1984), and Woodcock (1990).

Researchers have proposed alternatives to factor analytic theories of abilities. These include “cluster analyses” theory by R. Tryon (1939) and “facet analyses” theory by Louis Guttman (1954; see FACET THEORY; RADEX THEORY). A high-dimensional orthogonal model based on factor analysis was developed as the STRUCTURE-OF-INTELLECT MODEL theory of J. P. GUILFORD (1956). In the 1970s and 1980s some notable statistical features have been added to factor analysis methods (Lawley & Maxwell, 1971; Joreskog & Sorbom, 1979). Entire books about factor analysis include those by Gorsuch (1983) and McDonald (1985).

FACTOR ANALYSIS AS A FUNCTIONAL MODEL

The foundation of factor analysis can be expressed as a relationship between variables in a familiar equation: Response = $F(\text{Stimulus})$, or Outcome = $F(\text{Input})$, or simply $Y = F(X)$, where F indicates some kind of functional relationship. Figure 1 presents this kind of functional relationship as it is used in several common statistical methods (from McArdle & Lehman, 1992).

Figure 1a is labeled the *ANOVA function*. The left-hand side portrays a functional relation between stimulus (labeled S) and response (labeled R). The stimulus

levels plotted together on the X-axis form the “independent” variable and the response variable on the Y-axis is the “dependent” variable. In this example we have defined three specific levels of stimulation—labeled *zero*, *one*, and *two* (to designate the difference between, say, “control,” “experimental,” and “double exposure” conditions). In the analysis-of-variance model the best fitting line through the data is equivalent to the average differences between the means of each stimulus condition, but the interval on the X axis is not fixed. In ANOVA, the groups may not be ordered, and the best line may not be straight.

Figure 1b, labeled *the regression function*, is the same as Figure 1a, except that here the stimulus is allowed to be distributed over a wider range of the ordered X values. If a set of X values are selected at random (from all possible values of X), then these analyses can be especially effective. Using regression we can deal with unknown treatment effects, such as in random drug doses or clinical trials, or with a variable that is almost continuous (i.e., age or height). The best fitting straight line through these data is indicated by a regression coefficient (or slope), and this is defined as the expected value of the change in the response Y because of a one unit change in the stimulus X.

The three subplots of Figure 1c are used to define the *factor analysis function*. In the first subplot, the data points indicate scores on two response variates (R_1 and R_2). We often summarize this response—response association as a correlation coefficient ($C_{1,2}$). In contrast, the second and third plots of Figure 1c are latent stimulus-to-response regression plots that illustrate a few key points.

The common factor score (S) is an unobserved independent variable for many dependent measurements (R_m). This factor score is possibly also correlated with other common factors as well as with other variables not directly used in the analysis. The factor score is a key component in the specification of the external validity, interpretation, and use of the common factors. In contrast, the factor loading (L_m) is the linear regression coefficient (or slope) for a measured variable (R_m) as a function of the unobserved common factor score (S). The loadings of all measured variables on the common factors are collectively called the *factor pattern*. The set of factor loadings is a key component used

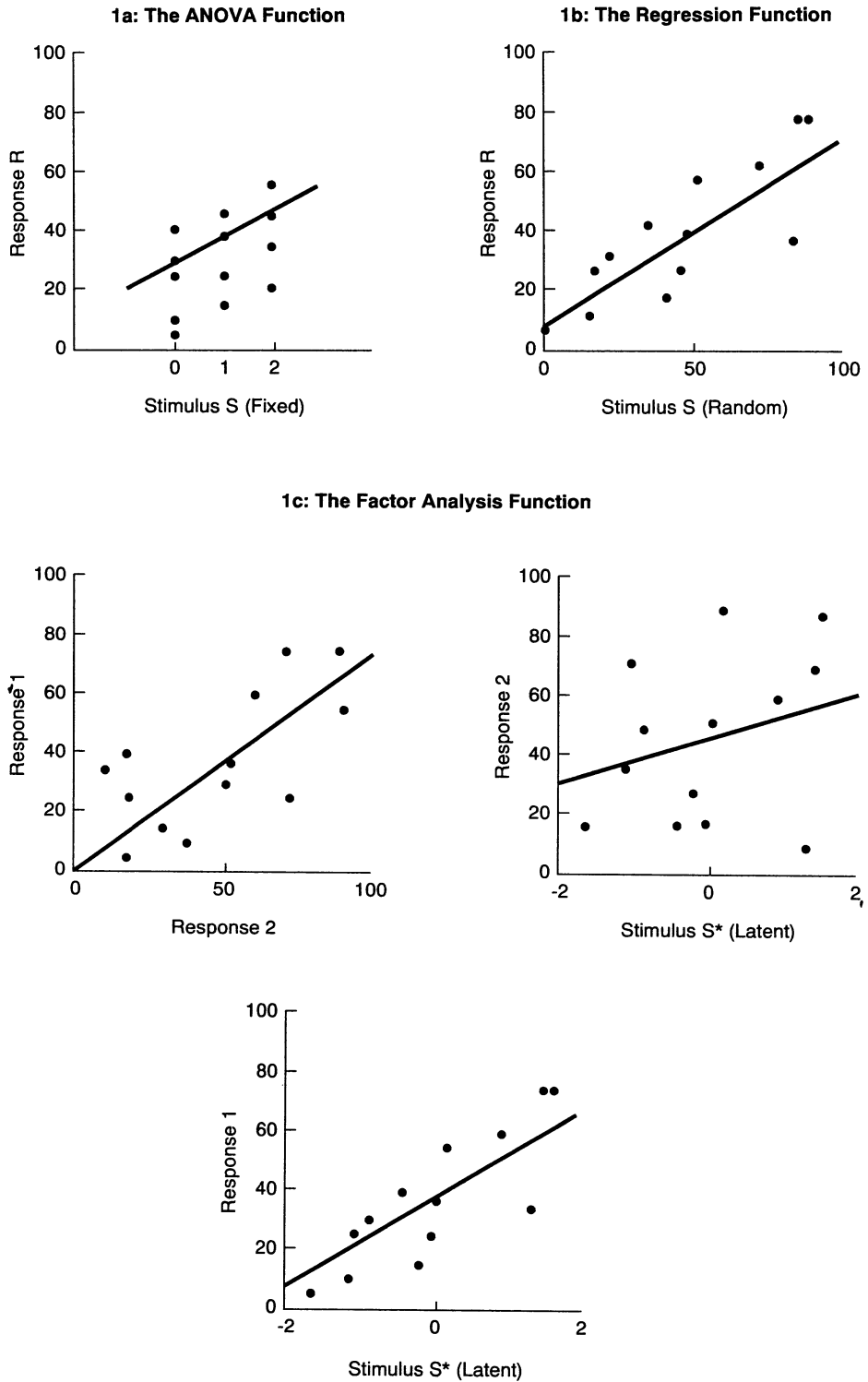


Figure 1
The functional form of several statistical models

in specifying the internal validity, interpretation, and meaning of the common factors.

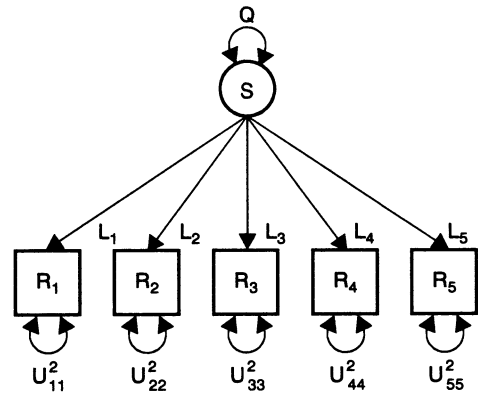
In factor analysis the stimulus-to-response functions (as in the last two plots in Figure 1c) are calculated from the response-with-response correlations (as in the first plot in Figure 1c). This indicates some practical differences between the three functional models of Figure 1. First, the factor function is mapped from one stimulus (the S s along the X-axis) to multiple response variables (several Y-axes), not one Y-axis. In practical work at least three responses are needed for each factor. Second, the independent variable (S) is not directly observed but is inferred from a broad pattern of influences. In contrast to ANOVA and regression functions, where X is merely a variable, the common factor represents the pattern, the configuration, the organizing principle, or the functional unity among many observed variables. In this way, the common factor, although not itself a construct, can represent the “evidence we have for the existence of a construct” (after Cattell, 1966).

ALTERNATIVE FACTOR ANALYSIS MODELS

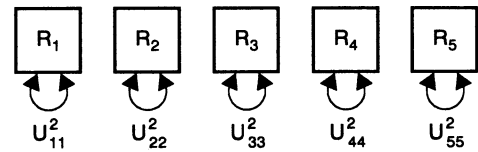
Factors in a factor analysis are ways to organize information. This organizing function is highlighted in the path diagrams (from Wright, 1921) of Figure 2. Observed variables (R_m) of the functional relations in Figure 1c are represented as squares, and latent factor scores (S) are here represented as circles. These path diagrams are a convenient way to express a wide variety of multidimensional mathematical models within a two-dimensional frame.

The model of Figure 2a is a path diagram from a single common factor. In theory, the common factor scores (S) determine, create, cause, or are the source of the correlations among the observed variables (R_m). Factor loadings (L_m) are drawn as one-headed arrows (from S to R_m), and these represent the strength of the relationship between the factor score and the observed score. The two-headed arrow attached to the common factor represents the variance of this factor. The two-headed arrows connected to the observed variables represent the variance of the unique factor for each separate measure.

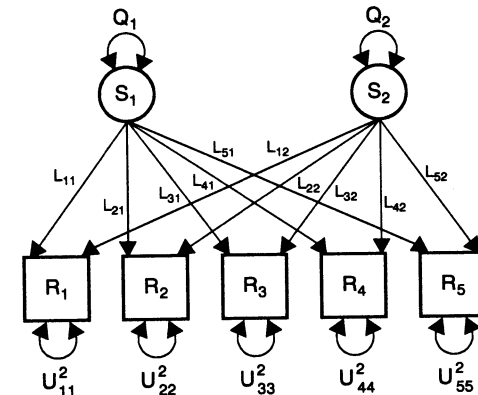
The model of Figure 2b shows an alternative to the model of Figure 2a: This is a “zero” common factor



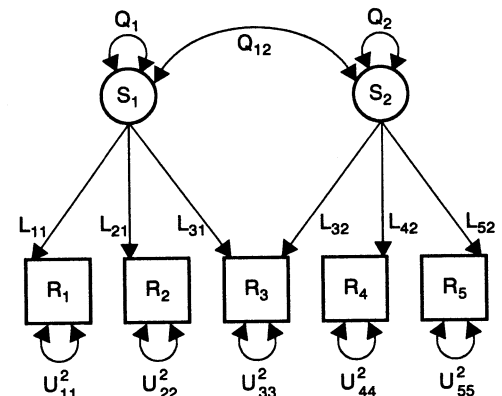
2a: The One Common Factor Model



2b: The Zero Common Factor Model



2c: The Orthogonal Two Common Factor Model



2d: The Oblique Two Common Factor Model

Figure 2

Path diagrams of alternative factor analysis models

model. In this model each variable has only unique variance of its own (symbolized by the two-headed arrows) and has no variance in common with other variables. Because no common factor (S) exists, the observations (R_m) are hypothesized to be completely independent of one another. That is, under this model the expected correlations between all variables are zero. This model is a rigorous way to represent the concept of independent psychological attributes, but it is usually far too extreme. Most reliably measured human abilities have been found to be correlated to some degree, as if one or more common factors are operating to produce responses.

The final two models of Figure 2 illustrate some further principles of multiple factor models. In these models two common factors (S_1 and S_2) influence the entire set of responses (R_m). In the model of Figure 2c the first and second influences are uncorrelated (they are not connected in this picture). This is labeled an *orthogonal score model* because the scores form an angle with a cosine of 90 degrees. In the model of Figure 2d a correlation (labeled $Q_{1,2}$) exists between the two factors. This is termed an *oblique score model* (here, based on a nonzero cosine). Both of these multiple factor alternatives are used in factor analysis experiments, but a strong case for the correlated factors model of Figure 2d has been made in research on human abilities (Thurstone, 1947; Meredith, 1965; McDonald, 1985; Cattell, 1987; Horn & McArdle, 1992).

FACTOR ANALYSIS CALCULATIONS

Factors in a factor analysis represent testable hypotheses about empirical data, and here calculations are important. These calculations are difficult by hand, but computing techniques, developed rapidly over the twentieth century, have now made them rather accessible and easy. There are several steps involved in a factor analysis, including: (1) data summary; (2) factor extraction; (3) factor rotation; and (4) factor score estimation. A wide variety of techniques exist within each of these steps, but only a summary of these procedures will be outlined here (Gorsuch, 1983; McDonald, 1985; Loehlin, 1992).

In most factor analytic studies a large number of individuals are measured on a smaller set of variables

in an effort to indicate a much smaller number of common factors. In a well-planned study, three or more variables are designed to indicate each theoretical factor. The data are initially organized into a data matrix of N rows (one for each individual or object) and M columns (one for each variable). In many studies, these scores are initially checked for outliers and influential observations, standardized, averaged over the rows (persons), and summarized in terms of the columns (variables). These procedures lead to a correlation matrix of product moment coefficients C_{ij} among the variables.

The next step in a factor analysis is termed the *factor extraction*. For a specified number of factors (K), an initial set of factor loadings is calculated from the correlations. This extraction is designed to yield statistically optimal factor loadings, similar to regression coefficients, using the model of proportionality described previously. Different optimization criteria can extract these factors, and popular extraction techniques are termed *maximum likelihood*, *principal axes*, and *principal components* (McDonald, 1985; McArdle, 1990). Each technique produces a unique set of values for a given number of factors (K). The communality (h_m^2) of the m -th measured variable is the variance explained by the K common factors, and communalities for all measured variables are estimated using one or more of the optimization techniques listed above. In contrast, the uniqueness of (u_m^2) of the m -th measured variable that is not in common with the other $m-1$ measured variables.

Decisions about the appropriate number of factors to retain are made by examining the results of a range of different factor extractions. One goal is to extract as much communality for each of the measured variables (M) by using a smaller number (K) of common factors. Historically, various characteristics of the variance of the initial latent variables (i.e., the latent roots or eigenvalues) were used to determine the appropriate number of factors; these techniques included the "root one" criterion, and the "scree test." It is common now to determine the appropriate number of common factors with a statistical test (see next section).

Another step is termed *factor rotation*. The entire set of K factor loadings for M variables can be transformed

mathematically or reorganized into different reference frames. Because the factor scores (S) are not directly observed, the factor loadings (L_m) can be seen from many different perspectives. Rotation does not affect the communalities of the variables or the overall fit of the model; it is primarily used as a way to test particular models or to simplify the substantive interpretation of the factors. This process is an important step in factor analysis because one kind of rotation of the common factors may appear to be more consistent with one theory than with another theory. A variety of formal rotational criteria have been developed to deal with this important problem.

L. L. THURSTONE (1931, 1947) developed a widely used rotational model termed *simple structure*. Thurstone stated these principles from ideas about the way natural mechanisms may emerge in the context of multiple variable measurement. He assumed that most causes affect only few outcomes and that most outcomes are affected by only a few causes. In accordance with this principle, simple structure rotation specifies a factor pattern with many zero loadings in each column (each factor affects only a few variables) and many zero loadings in each row (each variable is affected by only a few factors). Thurstone also knew that if simple structure could be achieved, then the interpretation of measured variables and common factors would be simplified.

Many formal mathematical and statistical rotation techniques have been devised to optimize these kinds of criteria. Some popular rotation techniques are based on orthogonal (uncorrelated) factors, such as Varimax or Equimax, and other techniques are usually based on oblique (correlated) factor rotations, such as Promax, Oblimin, Rotoplot, and Procrustes. Rotation theory remains an active area of research on factor analysis. For example, Cattell (1966) argued that factor invariance or equality of factor loadings across groups is the most important rotational goal (see Meredith, 1965; Horn & McArdle, 1992).

A further step is factor score estimation. Given K common factors, the task is to estimate the unobserved score (S_n) for each person and use these scores in further analyses. There are many techniques for calculating these scores, including least squares, Bartlett, and unity weighting procedures. However, since any

common factor model always implies unobserved scores for both K common and M unique factors, there are always more unobserved scores ($K + M$) than observed scores (M). For this reason, common factor scores can never be estimated exactly, and they reflect another source of indeterminacy in the common factor model (McDonald, 1985; McArdle, 1990). Several additional models have been developed to avoid such calculations, including factor extension analysis (Horn, 1976) and structural modeling (Loehlin, 1992).

TESTING A FACTOR ANALYSIS HYPOTHESIS

What are testable features of the common factor model? There are several ways to define the factor hypothesis or proportionality (e.g., with geometric relations, Figure 1c). In most contemporary treatments the factor model is evaluated statistically by the use of a goodness-of-fit test. The factor model stipulates that the expected correlation among any pairs of variables should be equal to the product of the estimated factor loadings (i.e., $E_{i,j} = L_i * L_j$). This expectation can then be compared to the observed correlations and a misfit (or model residual) can be obtained (i.e., $R_{i,j} = C_{i,j} - E_{i,j}$). Many different numerical indices of goodness-of-fit are based on the relative size of the model residuals.

In maximum likelihood estimation, the misfits are summarized into a likelihood ratio and an associated chi-square test. The test is used to answer the question: "Are the observed data consistent with the hypothetical model?" If the chi-square is small relative to its degrees of freedom, the model expectations are considered close to the observations; therefore, the model is considered to be plausible. If the chi-square is large relative to the degrees of freedom, then the model is not close to the data and rejected. A good fit does not imply that this K -factor model is the only model that fits these data, but a poor fit suggests that the model is not correct, at least in some respect. A great deal of research on model testing emphasizes the tradeoff between accuracy (good fit) and parsimony (having a smallest set of parameters; Browne & Cudeck, 1993). There are also techniques for calculating standard errors for the rotated factor loadings, so similar kinds of hypotheses testing can also be carried out

with specific loadings with a larger model. Statistical power analyses are now typical, and special considerations are made for binary and item-level data (McDonald, 1985). In all of these ways, a common factor model can be a clearly falsifiable set of statistical hypotheses.

The extraction and rotation steps described above are often called *exploratory factor analyses*. This approach is often distinguished from a confirmatory factor analysis where we represent a common factor hypothesis by placing a set of *a priori* restrictions on the factor loadings (or other model parameters) (McDonald, 1985). In confirmatory factor analyses, the factor extraction step and the factor rotation step are combined in one step termed *model estimation*.

The main benefits of confirmatory factor analyses are (1) We can directly represent and test specific factor analytic hypotheses. (2) We attempt to avoid some indeterminacy problems of the rotation step by providing a large number of model constraints. (3) We attempt to avoid the inherent indeterminacy of the factor score estimation step by making direct tests of hypotheses about common factor scores as part of a linear structural equation model (LISREL; Joreskog & Sorbom, 1979; Loehlin, 1992). One difficulty with this confirmatory approach is that we usually must have an extremely good idea about the correct model before we obtain the needed data and test the model fit. In practical work, exploratory and confirmatory techniques are best seen as two extremes on a research continuum (Nesselroade & Baltes, 1984; Nesselroade & Cattell, 1988).

ISSUES IN FACTOR ANALYSIS

Factors in factor analysis reflect information about empirical data. The kind of data chosen will determine how we calculate and interpret the factors. A wide variety of data collection designs can be used in factor analytic investigations (Cattell, 1952; Nesselroade & Cattell, 1988):

1. Factor analyses of data from item level measurement can be used with item response theory (IRT) analyses (LATENT TRAIT THEORY) or multidimensional scaling (MDS) to test ideas about the *basic measurement* of intelligence (McDonald, 1985).

2. Factor analyses of data from multiple group studies can be used in equivalence of measurement and

factorial invariance studies, and these studies test ideas about *quantitative versus qualitative* representations of intelligence (Meredith, 1965; Horn & McArdle, 1992).

3. Factor analyses of data from longitudinal studies can yield common factors, which reflect *aspects of development or processes* and the formal separation of stability from reliability (Horn, 1972; Nesselroade, 1972; McArdle & Anderson, 1990).

4. Factor analyses of data from studies of twins and families can yield information and the *genetic and non-genetic* factors of intellectual abilities (McArdle & Goldsmith, 1990; Loehlin, 1992).

5. Factor analyses of factor scores themselves may even be used to yield *higher-order components* of intellectual abilities (Cattell, 1971; Hakstian & Cattell, 1978).

The contribution of factor models of intellectual ability are clearly summarized by both John L. Horn (1988) and John B. Carroll (1993). An important message of the studies done to date is that a single common factor does not account for all the observed variation within the matrix of relationships. The results of a large number of studies indicate this same conclusion. The scientific evidence offered by factor analysis seems to indicate the humans have multiple factors of intellect and these change in different ways over age and circumstance. Yet, in the midst of this evidence, some researchers still view the single factor *g* model as a good representation of all things intellectual, and it is still common practice to add up different variable scores to obtain a single IQ score (Wechsler, 1981). Many factor analytic results, however, suggest that if only one IQ score is used, then important aspects of intellectual diversity are missed. As such, we need to ask why *g* and IQ continue to be popular concepts (Snyderman & Rothman, 1990).

Factor analysis is possibly the only widely used statistical method that has been derived from substantive problems in psychology. The concept of a factor as a "regularity in data" is a useful principle for all data analysis and all experimental design. Good factor analytic studies require representative collections of subjects and variables, a thoughtful use of automatic computerized procedures, and a broad awareness of alternative models. The testable hypotheses of factor analysis do not necessarily fit any real data, and we use factor analysis to check theory against the reality of

factual information. The factor model is a scientifically meaningful and functionally useful way to avoid the unbridled advocacy of our own pet theories. The factor analysis model, when precise in its testing implications and based on sound principles of contemporary data analysis, often offers a good scientific fit to our psychological theory.

BIBLIOGRAPHY

- BORING, E. G. (1950). *A history of experimental psychology*. New York: Appleton.
- BURT, C. L. (1909). Experimental tests of general intelligence. *British Journal of Psychology*, 3, 94–177.
- BROWNE, M., & CUDECK, R. (1993). Alternative ways of assessing model fit. In K. Bollen & S. Long (Eds.), *Testing structural equation models* (pp. 136–162). Beverly Hills, CA: Sage Publications.
- CARROLL, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- CATTELL, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, 38, 592.
- CATTELL, R. B. (1952). The three basic factor-analytic research designs—their interrelations and derivatives. *Psychological Bulletin*, 49, 499–520.
- CATTELL, R. B. (1966). *Handbook of multivariate experimental psychology*. New York: Rand McNally.
- CATTELL, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton-Mifflin (revised 1987 by North-Holland).
- CATTELL, R. B. (1987). *The scientific use of factor analysis in the behavioral and life sciences*. New York: Plenum.
- GORSUCH, R. (1983). *Factor analysis* (2nd ed.) Hillsdale, NJ: Erlbaum.
- GOULD, S. J. (1981). *The mismeasure of man*. New York: W. W. Norton.
- GUILFORD, J. P. (1956). The structure of intellect. *Psychological Bulletin*, 53, 267–293.
- GUSTAFSSON, J.-E. (1984). A unifying model for intellectual abilities. *Intelligence*, 8, 179–203.
- GUTTMAN, L. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 258–348). Glencoe, IL: Free Press.
- HAKSTIAN, A. R., & CATTELL, R. B. (1978). Higher-stratum ability structures on the basis of twenty primary abilities. *Journal of Educational Psychology*, 70, 657–669.
- HORN, J. L. (1972). State, trait and change dimensions of intelligence. *British Journal of Educational Psychology*, 4(2), 159–185.
- HORN, J. L. (1976). On extension analysis and its relation to correlations between variables and factor scores. *Multivariate Behavioral Research*, 11, 320–331.
- HORN, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade & R. B. Cattell (Eds.), *The handbook of multivariate experimental psychology*. (Vol. 2, pp. 645–685). New York: Plenum.
- HORN, J. L., & CATTELL, R. B. (1966). Age differences in primary mental ability factors. *Journal of Gerontology*, 21, 210–220.
- HORN, J. L., & ENGSTROM, R. O. (1979). Cattell's scree test in relation to Bartlett's chi-square test and other observations on the number of factors problem. *Multivariate Behavioral Research*, 14, 283–300.
- HORN, J. L., & MCARDLE, J. J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research*, 18(3), 117–144.
- JENSEN, A. R. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures*, 7, 93–118.
- JORESBOG, K. G., & SORBOM, D. (1979). *Advances in factor analysis and structural equation models*. Cambridge, MA: Abt Books.
- LAWLEY, D. N., & MAXWELL, A. E. (1971). *Factor analysis as a statistical method*. London: Butterworths.
- LOEHLIN, J. C. (1992). *Latent variable models: An introduction to factor, path and structural analysis*. Hillsdale, NJ: Erlbaum.
- MCARDLE, J. J. (1990). Principals versus principles in factor analysis. *Multivariate Behavioral Research*, 25(1), 81–87.
- MCARDLE, J. J., & ANDERSON, E. (1990). Latent variable growth models in research on aging. In J. E. Birren & K. W. Schaie (Eds.), *The handbook of the psychology of aging* (pp. 21–44). New York: Plenum.
- MCARDLE, J. J., & GOLDSMITH, H. H. (1990). Some alternative structural equation models for multivariate biometric analyses. *Behavioral Genetics*, 20(5), 569–608.
- MCARDLE, J. J., & LEHMAN, R. S. (1992). A functionalist view of factor analysis. In D. F. Owens & M. Wagner (Eds.), *Progress in modern psychology: The contributions of functionalism to modern psychology* (p. 167–187).
- MCDONALD, R. P. (1985). *Factor analysis and related methods*. Hillsdale, NJ: Erlbaum.
- MEREDITH, W. (1965). A method for studying differences between groups. *Psychometrika*, 30(1), 15–29.

- NESSELROADE, J. R. (1972). Note on the "longitudinal factor analysis" model. *Psychometrika*, 37(2), 187–191.
- NESSELROADE, J. R., & BALTES, P. B. (1984). From traditional factor analysis to structural-causal modeling in developmental research. In V. Sarris & A. Parducci (Eds.), *Experimental psychology in the future*. Hillsdale, NJ: Erlbaum.
- NESSELROADE, J. R., & CATTELL, R. B. (1988). *Handbook of multivariate experimental psychology*. (2nd ed.) New York: Plenum.
- SNYDERMAN, M., & ROTHMAN, S. (1990). *The IQ controversy, the media and public policy*. New Brunswick, NJ: Transaction Books.
- SPEARMAN, C. (1904). General intelligence objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- SPEARMAN, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.
- STERNBERG, R. J. (1985). Human intelligence: The model is the message. *Science*, 230(4730), 1111–1118.
- THURSTONE, L. L. (1931). Multiple factor analysis. *Psychological Review*, 38, 406–427.
- THURSTONE, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- TRYON, R. C. (1939). *Cluster analysis: Correlation profile or orthometric (factor) analysis for the isolation of unities in mind and personality*. Ann Arbor, MI: Edward Bros.
- WECHSLER, D. (1981). *Manual for the Wechsler Adult Intelligence Scale—Revised*. New York: The Psychological Corporation.
- WOODCOCK, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psycho-Educational Assessment*, 8, 231–258.
- WRIGHT, S. (1921). Correlation and causation. *Journal of Agricultural Research*, 20, 557–585.

JOHN J. MCARDLE

FACTOR ANALYSIS OF THE WECHSLER AND STANFORD-BINET SCALES The construct validity of a test is the extent to which the test measures a theoretical construct or trait. Factor analysis, a statistical procedure that helps identify the constructs that underlie a test, is one acceptable method

of demonstrating a test's construct validity. If a test provides scores on several scales, the presumption is that each scale corresponds to a meaningful theoretical construct. If factor analysis yields one factor to correspond to each scale, then such statistical results support the construct validity of the test. If the factors do not correspond to the scales, then the scales may not be measuring the constructs they were intended to measure. Consequently, the scores yielded by the scales may not be meaningful in a theoretical sense. The question of the factor analysis of Wechsler's scales and the Stanford-Binet, therefore, addresses the theoretical issue of whether these tests have construct validity and the practical issue of whether the obtained scores on these tests can be meaningfully interpreted.

In 1939, David WECHSLER published his first intelligence scale, the Wechsler-Bellevue, while working at Bellevue Hospital in New York City. Wechsler subsequently produced a series of scales that cover different portions of the age range between 3 and 74 years. These scales have undergone revision and restandardization; the currently used set of scales includes the 1989 Wechsler Preschool and Primary Scale of Intelligence—Revised (WPPSI-R) for ages 3 to 7; the 1991 Wechsler Intelligence Scale for Children—Third Edition (WISC-III) for ages 6 to 16; and the 1981 Wechsler Adult Intelligence Scale—Revised (WAIS-R) for ages 16 to 74. All Wechsler scales have one important feature in common: They are composed of equal, or nearly equal, numbers of verbal and performance (nonverbal) subtests, which are organized in separate scales, and they yield three intelligence quotients (IQs)—Verbal, Performance, and Full-Scale. The separate Verbal and Performance IQs correspond to the groupings of verbal and nonverbal subtests, respectively, and they are presumed to reflect "real" underlying abilities of children, adolescents, and adults. That presumption can be tested by the statistical technique of factor analysis.

If the abilities exist, then age-by-age factor analyses of Wechsler's intelligence scales should consistently produce two major factors, one from the five or six subtests that constitute the Verbal Scale and the other from the five or more subtests that constitute the Performance Scale. If the factor analyses produce markedly different factor structures, suggesting an alternate

organization of Wechsler's subtests, then the Verbal IQ and Performance IQ would lack factorial validity, a form of construct validity.

The Stanford-Binet Intelligence Scale was first published in 1916 by Lewis TERMAN of Stanford University. The scale was based on a French test devised by Alfred Binet and Théophile Simon in 1905. Composed of different mental tasks for different age levels between 2 years and "superior adult," it yielded a single IQ score. Subsequent editions of the Stanford-Binet were published by Terman and Maud Merrill between 1937 and 1972 and followed the same format. This one-score approach made factor analysis of limited or no value for assessing the test's factorial validity, and the inclusion of various mental tasks for the different age groups made it impractical to conduct such analyses.

In 1986, the Stanford-Binet-Fourth Edition (SB IV) was produced by a new team of authors, Thorndike, Hagen, and Sattler. They constructed the test by following a variant of the Cattell-Horn fluid-crystallized theoretical model; adopted Wechsler's subtest approach to test construction; and organized the test into four separate scales, each yielding a standard score known as a Standard Age Score (SAS). The four scales, which together yield a Composite SAS, are the Verbal Reasoning, Quantitative Reasoning, Abstract/Visual Reasoning, and Short-Term Memory scales. As is true for Wechsler's scales, the technique of factor analysis provides an empirical method for validating the meaningfulness of the four scales that constitute the SB IV.

Although factor analysis is a statistical procedure, there are many variations of the technique. No single criterion for determining the appropriate number of factors to interpret can be thought of as "right," and no single approach to factor analysis can be unambiguously defended as the "best" for any given test. Consequently, several alternative factor structure models, all of them reasonable, exist for any given test—and the answer to the question of whether factor analysis supports the test's construct validity may depend on which model one opts for. Usually considerable controversy and differing opinions occur about a test's factor structure: That has certainly been the case for Wechsler's scales—from the 1939 Wechsler-Bellevue to the 1991 WISC-III—as well as for the SB IV.

THE WECHSLER SCALES

Researchers have been conducting factor analyses of Wechsler's scales since the early 1950s; they have published hundreds of articles on the topic, using a plethora of techniques and subject groups. Jacob Cohen, who pioneered the early analyses of the Wechsler-Bellevue, WISC, and WAIS with a series of studies in the 1950s, argued that Wechsler's scales were composed of more than the two hypothesized factors, perhaps as many as five. Arthur Silverstein, whose work in the 1970s and 1980s greatly influenced the interpretation of Wechsler's factors, often insisted that two factors provide the best "fit" to the data. Alan Kaufman supported the notion that three factors usually offered the best explanation of the abilities that underlie most of Wechsler's tests. Some researchers have argued that there is only one meaningful Wechsler factor—a dimension of general intelligence.

Cohen's work, though important historically, represents overfactoring by contemporary standards. Some of his factors included just one subtest, while others represented an artificial splitting of the verbal dimension. Similarly, those who argue for just a single Wechsler factor reflect a minority, who attend exclusively to psychometric variables, ignoring the practical and clinical variables that govern appropriate test use and interpretation. Most researchers agree that Wechsler's scales are composed of either two or three factors, but there is some dispute (vehement on occasion) as to whether the "true" number is two or three.

When two factors are interpreted for Wechsler's various scales, they usually conform fairly closely to Wechsler's armchair assignment of subtests to either the Verbal Scale or the Performance Scale. These two factors are often called Verbal Comprehension and Perceptual Organization. The subtests that are most likely to be associated with neither scale, or with the "wrong" scale, are Digit Span, Coding (called Digit Symbol on the adult scales), and Picture Arrangement (putting pictures in the right order to tell a story). Digit Span is repeating a series of digits both forwards and backwards. It is a supplementary Verbal Scale subtest on the WISC-III (and on the WISC and WISC-R before it), and does not contribute to a child's Verbal

IQ; therefore, its failure to be associated with most Verbal Scale factors does not detract from the construct validity of Verbal IQ for children. Coding/Digit Symbol entails rapidly copying symbols that are paired with digits. It is more of a clerical than a mental task and does not seem to measure the kind of nonverbal thinking that the other Performance Scale subtests assess. Although it is on the Performance Scale, Picture Arrangement seems to require a good amount of verbal mediation to interpret and arrange the pictures.

When three factors are interpreted for children or adults, the third factor is usually composed of the Arithmetic (solving oral arithmetic problems) and Digit Span subtests, but sometimes includes Coding/Digit Symbol as well. Arithmetic, included on the Verbal Scale, is often associated with both the Verbal Comprehension factor and the third factor. The meaning of the third factor has been open to much dispute; it was named Freedom from Distractibility by Cohen, because the three subtests are considered to be the ones most likely to be susceptible to a child's or adult's distractible behavior. Even Cohen vacillated, however, and sometimes named the factor Memory. Others have called the factor Number Ability, Sequential Ability, Attention-Concentration, and Sequential Processing.

Kaufman has suggested that the factor means different things for different examinees; he has advised examiners to interpret the factor only when the person performs consistently on the component subtests and when the person's level of performance on the factor differs significantly from performance on at least one of the other two factors. He further advises that the meaning of the third factor for a given person can only be interpreted in the context of that person's background information, observed behaviors during the testing session, and test scores on other pertinent Wechsler subtests and items.

Those who have argued against interpreting the third factor have noted that it does not always emerge when strict empirical procedures are applied; it does not emerge for some ethnic groups (e.g., Native Americans) or clinical populations; it is much smaller in magnitude than the Verbal Comprehension and Perceptual Organization factors; and its subtest composition frequently varies from sample to sample. Those who defend the third factor note that (1) a Digit Span/

Arithmetic factor—either alone or joined by another subtest or two—has emerged for every school-age and adult Wechsler test since the Wechsler-Bellevue and (2) the existence of this factor has been confirmed for normal individuals of both genders and most ethnic groups, as well as for a wide variety of clinical, educationally handicapped, and neurologically impaired populations. For example, clearly identifiable distractibility factors have emerged on the WISC-R for separate groups of male, female, white, Hispanic, learning-disabled, psychiatrically impaired, and mentally retarded children.

Whether one interprets two or three factors for the WISC, WISC-R, WISC-III, WAIS, WAIS-R, and the two forms of the Wechsler-Bellevue, the construct validity of Wechsler's scales has been given strong support. The Verbal Comprehension and Perceptual Organization factors that emerge in isolation or accompanied by a third factor are each large and robust. Furthermore, the four highest-loading subtests on the Verbal Comprehension factor are invariably members of the Verbal Scale. These are Information (answering general information questions), Similarities (telling how two verbal concepts are alike), Vocabulary (defining words), and Comprehension (answering socially relevant questions). Arithmetic is often the fifth highest, and the highest-loading four or five subtests on the Perceptual Organization factor are commonly Performance Scale subtests. In addition, substantial loadings on the first unrotated factor by most Wechsler subtests (g loadings, where g is the general factor) support the combination of the Verbal and Performance subtests to yield a Full-Scale IQ.

The factor structure of the newest Wechsler test, the WISC-III, differs from that of all previous tests because of the addition of a thirteenth subtest, a supplementary Performance Scale subtest called Symbol Search.

WPPSI-R. The WPPSI-R, like the WPPSI before it, has a clear-cut two-factor structure, which conforms quite closely to Wechsler's division of subtests into the Verbal and Performance scales. Unlike factor analyses for the Wechsler tests designed for school-age children and adults, a distractibility factor does not emerge for the WPPSI-R, nor does any other meaningful third factor. For the WPPSI-R, there are just two slight exceptions to the unambiguous split of

subtests into Verbal and Performance dimensions: Picture Completion (finding the missing part of a picture) is an equal measure of Verbal and Performance ability for 3- to 4-year-olds (perhaps because children usually respond to this Performance subtest by naming the missing part of a picture); and Arithmetic is associated with the Performance as well as the Verbal factor for all ages (perhaps because this Verbal subtest includes numerous visual stimuli).

WISC-III. The inclusion of the Symbol Search (rapidly scanning a row of symbols to find target symbols) subtest in the WISC-III has altered the factor structure of the WISC-III compared to the structure of all other Wechsler tests. Verbal Comprehension and Perceptual Organization factors emerge as usual, and these staples are accompanied by a two-subtest distractibility factor (Arithmetic and Digit Span) and a two-subtest Processing Speed factor (Coding and Symbol Search). The WISC-III test manual provides extensive support for the four factors based on data on both normal and exceptional (gifted and cognitively challenged) children, and the WISC-III provides examiners with tables for computing a supplementary set of standard scores (Factor Indexes) on the four factors. Some investigators, such as Jerome Sattler, however, have argued instead for a three-factor solution, insisting that the distractibility factor is not meaningful.

WAIS-R. The evidence from a diversity of normal and clinical samples supports the WAIS-R as having a three-factor structure. The Verbal Comprehension factor is typically composed of all Verbal subtests except Digit Span; the Perceptual Organization factor is defined primarily by Block Design (constructing abstract designs out of blocks), Object Assembly (putting together cut-up picture puzzles), and Picture Completion; and the distractibility factor is composed of Arithmetic, Digit Span, and sometimes Digit Symbol. Picture Completion also has an association with the Verbal Comprehension factor, and Picture Arrangement relates about equally to the Verbal and Performance dimensions. Digit Symbol loads on the Perceptual Organization factor, rather than the distractibility factor, for some clinical samples (e.g., brain-damaged and neuropsychiatric patients). Clear-cut distractibility factors have emerged on the WAIS-R for separate groups of adult males, females, whites, blacks, psychiatric inpatients, medical patients, learning-dis-

abled individuals, brain-damaged patients, and neuropsychiatric patients.

STANFORD-BINET, FOURTH EDITION

The Technical Manual for the SB IV attempts to offer support for the test's four factors using a variant of the technique of confirmatory analysis. Jerome Sattler, the test's third author, proposes alternate solutions for the SB IV: For ages 2 to 6, he suggests two factors (Verbal Comprehension and Nonverbal Reasoning/Visualization); for ages 7 to 23, he proposes three factors (Verbal Comprehension, Nonverbal Reasoning/Visualization, Memory). Robert M. Thorndike, son of the SB IV's first author, disagrees with aspects of the methodology used by Sattler, used in the SB IV Technical Manual, and used by other investigators of the SB IV's factor structure. His reanalysis of the data by both exploratory and confirmatory techniques, however, is largely in agreement with previous analyses. Thorndike's results conform closely with Sattler's, and his evaluation of the SB IV by confirmatory techniques provides only weak support for a four-factor structure. In particular, a Quantitative Reasoning factor usually fails to emerge; when it does, it is weak. Large *g* factors in the unrotated solutions do, however, support the construct validity of the SB IV's global Composite SAS.

The bulk of research on the SB IV's factor structure does not support the construct validity of the SB IV; it disputes the interpretation of the four standard scores provided for the separate scales. The SB IV is a two-factor test for preschool and primary-age children (as is the WPPSI-R) and conforms to Wechsler's verbal-nonverbal dichotomy. For ages 7 through early adulthood, the verbal and nonverbal factors are joined by a memory factor—just as Wechsler's Verbal Comprehension and Perceptual Organization factors are joined by Freedom from Distractibility (sometimes called Memory) for most school-age and adult samples. The SB IV analyses, therefore, suggest that the test's structure conforms more to the Wechsler model, and to the results of factor analyses of Wechsler's tests, than to the four-scale structure that defines the new Binet.

Regarding test interpretation for ages 2 to 6, R.M. Thorndike recommends grouping the subtests as ver-

bal versus nonverbal, and interpreting large differences between the two areas; primary focus, however, should be on the Composite SAS. For ages 7 and above, he notes that the composition of the Verbal and Memory factors corresponds fairly closely to the Verbal Reasoning and Short-Term Memory scales, respectively; the Abstract/Visual factor, however, blends tasks from the Quantitative Reasoning Scale with spatial/reasoning subtests. Empirical support exists for interpreting the Verbal Reasoning, Short-Term Memory, and Composite SAS's; the Abstract/Visual Reasoning and Quantitative Reasoning SAS's may not correspond to meaningful, unitary dimensions of intelligence.

CONCLUSIONS

Whether one accepts two-, three-, or four-factor solutions for Wechsler's various scales, the results of both exploratory and confirmatory factor analyses are in agreement: Wechsler's three IQs, including the separate Verbal IQ and Performance IQ, have high construct validity, and may be interpreted in the way the test author intended. In contrast, exploratory and confirmatory factor analyses of the SB IV are neither consistent nor unified, and differences occur across the age range. Consequently, the construct validity of the SB IV is only partially supported by factor analysis, and the proper way to interpret a person's test performance on the SB IV remains a controversial issue. At best, the four separate SAS's yielded by the SB IV require cautious interpretation because the scales may not all correspond to meaningful theoretical constructs. Nevertheless, it must be remembered that factor analysis is only one type of evidence for construct validity, and other types may be more supportive of the SB IV's separate scales.

(See also: WECHSLER SCALES OF INTELLIGENCE.)

BIBLIOGRAPHY

- BRACKEN, B. A., & FAGAN, T. K. (EDS.). (1990). Special Issue—Intelligence: Theories and practice. *Journal of Psychoeducational Assessment*, 8(3).
- KAUFMAN, A. S. (1979). *Intelligent testing with the WISC-R*. New York: Wiley.
- KAUFMAN, A. S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn & Bacon.

- KAUFMAN, A. S., HARRISON, P. L., & ITTENBACH, R. F. (1990). Intelligence testing in the schools. In T. B. Gutkin & C. R. Reynolds (Eds.), *The handbook of school psychology* (2nd ed.) (pp. 289–327). New York: Wiley.
- LECKLITER, I. N., MATARAZZO, J. D., & SILVERSTEIN, A. B. (1986). A literature review of factor analytic studies of the WAIS-R. *Journal of Clinical Psychology*, 42, 332–342.
- MATARAZZO, J. D. (1972). *Wechsler's measurement and appraisal of adult intelligence* (5th ed.). New York: Oxford University Press.
- SATTLER, J. M. (1988). *Assessment of children* (3rd ed.). San Diego, CA: Author.
- SATTLER, J. M. (1992). *Assessment of children: WISC-III and WPPSI-R supplement*. San Diego, CA: Author.
- THORNDIKE, R. L., HAGEN, E. P., & SATTLER, J. M. (1986). *Technical manual—Stanford-Binet Intelligence Scale—Fourth Edition*. Chicago: Riverside.
- WECHSLER, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence—Revised (WPPSI-R)*. San Antonio, TX: The Psychological Corporation.
- WECHSLER, D. (1991). *Manual for the Wechsler Intelligence Scale for Children—Third Edition (WISC-III)*. San Antonio, TX: The Psychological Corporation.

ALAN S. KAUFMAN

FAMILY ENVIRONMENTS A big question in the nature versus nurture debate about intelligence is the relative influence of family environments versus heredity on the development of children's intelligence. Social scientists have long regarded the family environment as a principal influence on children's intelligence. The remarkably strong parent-child resemblance for intelligence quotient (IQ) was often cited as proof of influence: Relatively bright parents tend to have bright children, whereas relatively dull parents tend to have intellectually dull children. Social scientists have also recognized a second possibility—that familial resemblance in IQ may stem from genes shared by parent and child, which may in turn affect IQ development through the physiology of the nervous system (Rowe, 1991).

Just as genetic differences can make one child tall and another short or one obese and another thin, they may make one child bright and another dull. Sometimes scientists employ the phrase "genes for intelligence." This phrase does not mean that a single "IQ

gene” determines whether a child is bright or dull—many “IQ-relevant genes” would be involved. A better interpretation of the phrase is that a large number of genetic differences can produce IQ variation.

Variation in IQ also may derive from environmental differences. Some environmental differences exist between families and may make children in one family different in IQ from children in another family. If these family environments are of a kind experienced in common by parents and children in a household or by brothers and sisters, the term *shared environmental effects* applies to them. By definition, such shared effects—such as exposure to books in the home or the general level of nutrition—operate to make reared-together family members similar in IQ but different from members of other families. In contrast, other environmental effects operate to make family members dissimilar in IQ; they are *nonshared environmental effects* or *within-family effects*. One sibling experiences oxygen deprivation at birth whereas another has an uneventful birth, or one sibling receives a chemistry set as a birthday present and another receives a violin.

Parents pass genes to their offspring and they also “construct” many aspects of family environment, such as intellectual stimulation. These dual pathways of influences create a confusion of genes and environmental influences in studies of nuclear families. For this reason, certain research designs are useful to separate the two kinds of influences. Such research takes advantage of the diversity of family types, both biologically and socially. For instance, the study of twins compares IQ outcomes across levels of genetic relatedness (identical versus fraternal twins); the adoption study compares biologically unrelated persons reared together. These research designs have led to a surprising discovery: Shared family environmental effects on intelligence are less than many social scientists would have believed. Furthermore, the effects may diminish progressively as children mature.

HERITABILITY OF IQ

The heritability coefficient for IQ represents the degree to which IQ variation derives from genetic differences. Since 1900, an enormous body of evidence has accumulated on kinship correlations for IQ. T. J. Bouchard and M. McGue’s 1981 update of L. Erlenmeyer-Kimling and L. F. Jarvik’s 1963 summary of

family studies of IQ, reported familial correlations for over 40,000 pairings of relatives from over 100 different studies. Review studies have analyzed the corpus of these data using mathematical models that estimate heritability and shared environmental effects. On the basis of the IQ correlations in Table 1, J. C. Loehlin (1989) estimated the heritability of IQ as 47–58 percent of total IQ variation. Also using Bouchard and McGue’s review, other investigators estimated IQ’s heritability at 51 percent (Chipuer, Rovine, & Plomin, 1990). The genetic effect in Table 1 is especially evident in the correlations of biological relatives reared apart. An IQ correlation between identical twins raised apart is a direct estimate of heritability; this estimate in Table 1—72 percent—is actually higher than Loehlin’s or H. M. Chipuer and colleagues’ estimates using all IQ correlations together. It may reflect gene x gene interaction effects that identical twins but not other relatives share. Bouchard and colleagues (1990) showed that the IQ similarity of twins reared apart is not the result of placement into similar adoptive homes or social contacts between twins. By 1993, among most social scientists, the heritability of intelligence was no longer in doubt (Synderman & Rothman, 1987). (See also HERITABILITY.)

The heritability of IQ implies that genetic inheritance is quite important for IQ variation. Genes cannot be ignored in theories of IQ correlates, such as social class or mental health. Higher social class has an association with higher IQ in offspring because bright children manage to rise in social status relative to their parents and give their own children the genes favorable to high IQ. Data supports this biological pathway of influence. Sons brighter than their fathers tend to rise above their fathers in social status, and vice versa for sons duller than their fathers (Waller, 1971).

SHARED ENVIRONMENTAL INFLUENCE ON IQ

Loehlin’s mathematical models also estimated the extent of shared environmental influence. Loehlin’s estimates were 39 percent for twins, 27 percent for nontwin siblings, and 22 percent for parent–child. Chipuer and colleagues’ models yielded estimates from 11 percent to 35 percent, also depending on which pairings of biological relatives were involved. In Table

TABLE 1
IQ correlations from Bouchard and McGue's summary (1981)

<i>Kinship</i>	<i>Mean Correlation</i>	<i>Number of Pairs</i>
Biologically unrelated persons reared together		
One adoptive sibling, one biological	.29	345
Two adoptive siblings	.34	369
Adoptive parent-adoptive child	.19	1,491
Biologically related persons reared together		
Siblings	.47	26,473
Parent-Child	.42	8,433
Fraternal twins	.60	5,533
Identical twins	.86	4,672
Biologically related persons reared apart		
Parent-Child	.24	720
Siblings	.24	203
Identical twins	.72	65

SOURCE: Adapted from Loehlin (1989) by permission.

1, shared family environmental effects that contribute to IQ similarity can be seen in the IQ correlations for unrelated persons reared together.

Nevertheless, some data place a caution on these estimates of shared environmental influence because they may be stronger in children than adults. Consider the IQ correlation for unrelated siblings reared together. Figure 1 shows the IQ correlations from nine studies of unrelated siblings in childhood and four studies of them in adulthood. The mean correlation in childhood was .25, suggesting that shared environmental effects constituted 25 percent of total IQ variation. In contrast, siblings who were raised together but tested as adults were unlike in IQ; their mean IQ correlation was about zero ($r = -.01$). Thus, by adulthood, variation in shared family environments no longer seems to influence intelligence directly.

This discovery requires confirmation from other research designs. Table 2 compares the IQ correlations for pairs of adult relatives raised together versus raised apart. Shared family environmental effects would appear as greater IQ resemblance in pairs of relatives reared together—where they share an experience of a

common family environment—than in pairs reared apart in different households. Contrary to this expectation, the correlations for reared-apart versus reared-together biological relatives were quite similar. For example, reared-together full siblings correlated only .07 more than reared-apart full siblings (.54 versus .47). The reared-together identical twins correlated only .11 more than those reared apart. This overall pattern leads to a conclusion that family-tied environmental influences have only small effects on the IQ level children attain as adults.

Thus childhood-rearing circumstances may not determine adult IQ (at least when the rearing levels fall in range from the working to professional social classes). If children reared in a single professional-class family are as different as children raised in a working-class and a professional-class family, then “swapping” children between these two kinds of families should leave adult IQs unaffected—a surprising notion for many social scientists steeped in strong beliefs about family environmental effects on IQ.

This conclusion, however, does not deny the importance of experience for achieving a high adult IQ—

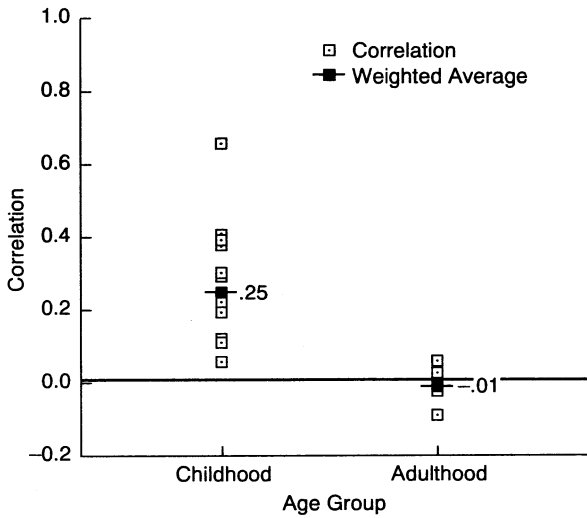


Figure 1
Weighted average correlation for unrelated siblings reared together in childhood and adulthood

extensive vocabularies, good problem-solving skills, and the ability to extract meaning from a difficult paragraph are not inborn traits. The conclusion means that even children who were raised in what may appear to be poor family environments can usually find those experiences needed to attain above-average IQs if they possess genes favorable to high IQ. For instance, in working-class homes bright adoptive children may find encouraging teachers, may read newspapers and books, and may have scholarly friends. From such per-

sonal experiences their intellects may grow, despite fewer “objective” signs of intellectual stimulation in their working-class adoptive homes than those in professional-class families. The effect of experience may increase with maturation because greater opportunities exist outside the family context as children grow older; in S. Scarr and K. McCartney’s (1983) phrase, genes may “drive” experience because people with different genetic dispositions actively find social environments more supportive of their IQ dispositions as they become more sophisticated about the world. J. H. Waller’s (1971) data on social mobility suggest one consequence of this “niche picking”—brighter children tend to move into more intellectually challenging occupations and duller children move into less challenging ones.

Because behavior genetic studies have included few families living in dire poverty, family environments of this latter type possibly have stronger influences on adult IQ than those implied in Table 2. The conclusion put forth here would extend to about 80 percent of the American public, not to those children in deprived or abusive home situations.

NONSHARED ENVIRONMENTAL EFFECTS ON IQ

In Table 1, the twin correlation for identical twins raised apart was .72—less than the statistical reliability of IQ tests (about .90). This gap between correlations

TABLE 2
IQ correlations on adult data from four kinship groups reared apart and together

<i>Kinship Group</i>	<i>IQ Correlation</i>	<i>Number of Pairs</i>	<i>Kinship Genetic Correlation</i>
Reared apart			
Half-siblings	.22	64	.25
Full siblings	.47	28	.50
Fraternal twins	.35	112	.50
Identical twins	.75	158	1.00
Reared together			
Half-siblings	No data		.25
Full siblings	.54	271	.50
Fraternal twins	.39	178	.50
Identical twins	.86	190	1.00

implies a third source of variation—the nonshared environment—which can make genetically identical twins differ in their true IQs. About 10–20 percent of the total IQ variation may derive from nonshared environmental influences, which make family members dissimilar.

D. C. Rowe and R. Plomin (1981) summarized a diverse set of influences that may contribute specifically to IQ differences:

1. accidental influences (perinatal traumas)
2. differential sibling interactions
3. family structure
4. differential parental treatments
5. extrafamilial influences, such as peers and specific teachers.

Of these influences, the best documented empirically is the teratogenic effects of alcohol and other substances on fetal development. Prenatal alcohol exposure can cause FETAL ALCOHOL SYNDROME, a distinct disorder with abnormalities of physical development and mental retardation as symptoms. Even modest levels of prenatal alcohol exposure may have subtle detrimental effects on cognitive functioning (Streissguth et al., 1989). Exposure to teratogenic substances is, however, also a potentially preventable source of lower IQs.

For many years, family structure (i.e., birth order) was thought to be a major nonshared source of IQ variation. The principal supporting data consisted of a regular decline of IQ with increasing birth order in aggregate cross-sectional data (Zajonc & Markus, 1975). The problem with this evidence is that parents with lower IQ have tended to have larger families (at least after the U.S. baby boom generation), which statistically confounds birth order with parental IQ. This confusion can be avoided by comparing siblings of different birth orders. Using this procedure, the relation of birth order and IQ diminishes and often disappears (Retherford & Sewell, 1991; Rodgers, 1984). Moreover, birth order fails to moderate the IQ similarity of siblings—first born children resemble second-borns in IQ as much as they resemble children born fourth (Rodgers & Rowe, 1985). Birth order has lost its luster as an explanation for IQ variation.

The family-tied explanations of IQ variation were supported for children but not for adults. Specific

work on them as nonshared environmental influences is, however, not abundant. Family environments do not appear to explain the dissimilarities in IQ between raised-apart identical twins (Bouchard, 1983; Bouchard et al., 1990) and thus may not act as nonshared environmental influences on IQ. Extrafamilial influences, of course, affect IQ—they are the experiences needed to nurture nature—but they may not be reproducible nonshared influences because their major effect may depend on inherited dispositions favorable or unfavorable toward IQ.

CONCLUSIONS

IQ is a substantially heritable trait, but dissimilar from nonintellectual personality traits as its earliest manifestations in childhood are also strongly influenced by the shared environment of rearing. In adulthood, the shared environmental effects diminish considerably, and IQ variation depends more on genetic inheritance and nonshared environmental effects. The picture is one of many children who, given the opportunities, actively approach a level of genetic potential by seeking experiences that reinforce their cognitive capacities. Children may veer away from intellectual tasks that are either too effortful or too unchallenging for them so that their genetic potential is eventually expressed.

(See also: BIRTH ORDER, SPACING, AND FAMILY SIZE; NATURE, NURTURE, AND DEVELOPMENT; PARENTING AND INTELLIGENCE.)

BIBLIOGRAPHY

- BOUCHARD, T. J., JR. (1983). Do environmental similarities explain the similarity in intelligence of identical twins reared apart? *Intelligence*, 7, 175–184.
- BOUCHARD, T. J., JR., LYKKEN, D. T., MCGUE, M., SEGAL, N. L., & TELLEGEN, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science*, 250, 223–228. This article reports the fascinating similarities of reared-apart identical twins but requires good knowledge of statistics.
- BOUCHARD, T. J., JR., & MCGUE, M. (1981). Familial studies of intelligence: A review. *Science*, 250, 223–238. This reference is the most current compilation of the world literature on IQ correlations.

- CHIPUER, H. M., ROVINE, M. J., & PLOMIN, R. (1990). LISREL modeling: Genetic and environmental influences on IQ revisited. *Intelligence, 14*, 11–29.
- ERLENMEYER-KIMLING, L., & JARVIK, L. F. (1963). Genetics and intelligence: A review. *Science, 142*, 1477–1479.
- LOEHLIN, J. C. (1989). Partitioning environmental and genetic contributions to behavioral development. *American Psychologist, 44*, 1285–1292.
- RETFERFORD, R. D., & SEWELL, W. H. (1991). Birth order and intelligence: Further tests of the confluence model. *American Sociological Review, 56*, 141–158.
- RODGERS, J. L. (1984). Confluence effects: Not here, not now! *Developmental Psychology, 20*, 321–331.
- RODGERS, J. L., & ROWE, D. C. (1985). Does contiguity breed similarity? A within-family analysis of nonshared sources of IQ differences between siblings. *Developmental Psychology, 21*, 743–746.
- ROWE, D. C. (1991). Heredity. In V. J. Derlega, B. A. Winstead, & W. H. Jones (Eds.), *Personality: Contemporary theory and research* (pp. 56–85). Chicago: Nelson-Hall. A readable introduction to the field of behavior genetics.
- ROWE, D. C., & PLOMIN, R. (1981). The importance of non-shared (E_i) environmental influences in behavioral development. *Developmental Psychology, 17*, 517–531.
- SCARR, S. (1993). Biological and cultural diversity: The legacy of Darwin for development. *Child Development, 64*, 1333–1353.
- SCARR, S., & MCCARTNEY, K. (1983). How people make their own environments: A theory of genotype→environment effects. *Child Development, 54*, 424–435.
- SNYDERMAN, M., & ROTHMAN, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist, 42*, 137–144.
- STREISSGUTH, A. P., BARR, H. M., SAMPSON, P. P., DARBY, B. L., & MARTIN, D. C. (1989). IQ at age 4 in relation to maternal alcohol use and smoking during pregnancy. *Developmental Psychology, 25*, 3–11.
- WALLER, J. H. (1971). Achievement and social mobility: Relationships among IQ score, education, and occupation in two generations. *Social Biology, 18*, 255–263.
- ZAJONC, R. B., & MARKUS, G. B. (1975). Birth order and intellectual development. *Psychological Review, 82*, 74–88.

DAVID C. ROWE
JOSEPH LEE RODGERS

FETAL ALCOHOL SYNDROME A large body of experimental research has established that alcohol is a teratogen—a substance that can cause defects in a fetus by affecting the growth and proper formation of its body and/or brain. Prenatal exposure to alcohol is, therefore, a preventable cause of birth defects. Abel and Sokol (1987) have highlighted exposure to alcohol before birth as one of the leading known causes of mental retardation in the Western world, and researchers have begun to define the full range of alcohol-related birth defects.

DEFINITIONS

The term *fetal alcohol syndrome* (FAS) was first used by Jones, Smith, and their colleagues (Jones & Smith, 1973; Jones et al., 1973) to describe a pattern of abnormalities seen in children born to alcoholic mothers. Prior investigators had also observed growth retardation and abnormalities in such children. Since then, FAS has been established as a birth defect and medical diagnosis. Standard defining criteria were proposed by Rosett (1980), were updated by Sokol and Clarren (1989), and are still being refined. These are (1) prenatal and/or postnatal growth retardation (weight and/or length or height below the 10th percentile) when corrected for gestational age; (2) central nervous system involvement, including signs of intellectual impairment, developmental delay, neurological abnormality, behavioral dysfunction or deficit, and skull or brain malformations; and (3) a characteristic facial appearance, which includes short palpebral fissures (eye openings), a thin upper lip, a flat and long midface, and a long and/or flat philtrum (two ridges between the nose and upper lip). Optimal times to diagnose FAS are between 8 months and puberty.

THE RANGE OF ALCOHOL-RELATED BIRTH DEFECTS

There is a broad spectrum of alcohol-related birth defects, ranging from the full fetal alcohol syndrome to incomplete features of FAS, including more subtle cognitive-behavioral deficits. The term *possible fetal alcohol effects* (possible FAE) is a clinical and medical research category sometimes used when there is a clear history of prenatal alcohol exposure, but physical char-

FAS See FETAL ALCOHOL SYNDROME.

acteristics are not sufficient to warrant a diagnosis of FAS. Use of the term is still being clarified.

Children who meet the criteria for FAS appear to be born primarily to those mothers who drink large amounts of alcohol during pregnancy. Yet researchers have found subtle and enduring neurobehavioral and morphological (formation and structure) deficits to be associated with more moderate levels of prenatal alcohol exposure in infants and children. Such developmental research studies have not shown consistent consequences caused by similar conditions of prenatal alcohol exposure. This may be due to differences in sample size, number of more heavily exposed offspring within a sample, age at follow-up, sensitivity of research instruments, and other factors in research design.

HOW PRENATAL ALCOHOL EFFECTS OCCUR AND THE QUESTION OF A SAFE DRINKING THRESHOLD

Like those of other teratogens, the effects of alcohol depend on the amount, timing, and conditions of exposure. The effects of alcohol on fetal brain and body development are not reversible, and functional deficits (e.g., intellectual impairment and behavior problems) appear to be more common outcomes than physical malformations or growth impairment. The many possible effects of prenatal exposure to alcohol are under intensive study by researchers such as Clarren et al. (1990), Riley (1990), West (1986), and others, using animal models or examining human nervous system structure and function. Researchers studying human development, such as Coles et al. (1991), Day et al. (1991), Fried and Watkinson (1990), Greene et al. (1990), and Streissguth (1992; Streissguth et al., 1991; Streissguth, Barr, & Sampson, 1990; Streissguth & Clarren, 1992; Streissguth & LaDue, 1987), are investigating the impact of alcohol exposure before birth within the context of environmental influences after birth.

Since a wide range of deficits results from various amounts of prenatal alcohol exposure, a 1991 publication of the National Institute for Alcoholism and Alcohol Abuse stated that it may not be reasonable to search for an overall "safe" drinking threshold for fetal risk. For the past decade, the surgeon general has recommended that women not drink during pregnancy.

EXTENT OF THE PROBLEM

There are methodological difficulties in doing research on the incidence of FAS, and more data are needed. Available databases provide either under- or overestimates of the problem. It is difficult to establish the frequency of a birth defect that is hard to diagnose in a newborn and that must be diagnosed by careful history and physical examination rather than by a laboratory test. Recently, Little et al. (1990) found a 100 percent failure rate to diagnose FAS in newborns born in a very large hospital obstetrical service. For the U.S. population overall, the National Institute of Alcoholism and Alcohol Abuse (1990) estimates the incidence of FAS to be 1 to 3 per 1,000 live births. In some populations where alcohol abuse among women is prevalent, there appears to be a much higher incidence of FAS. In other studies, which may lack adequate examination and follow-up, there have been reports of a lower incidence of FAS.

PATTERN OF DEVELOPMENT IN INDIVIDUALS WITH FAS/FAE ACROSS THE LIFESPAN

Researchers are studying the physical and behavioral patterns of development in individuals with FAS, and some with possible FAE, from infancy through adolescence and into adulthood. Manifestations of the characteristic growth deficiency and facial anomalies change somewhat after puberty. Short stature and microcephaly (small head size) are the most notable growth deficiencies in adolescents and adults with FAS/FAE. Facial anomalies seem to be more subtle after puberty, with relatively short palpebral fissures, a smooth philtrum, and a thin upper lip still important discriminating features.

There are lasting cognitive and behavior problems among individuals with FAS/FAE, given their characteristic central nervous system involvement. Useful articles describing these lifespan problems are those by Giunta and Streissguth (1988) and by Streissguth et al. (1991). Studies indicate that in children diagnosed with FAS, typical deficits include problems in attention, memory, learning, and problem solving, as well as motor incoordination, impulsivity, and hyperactivity. A limited number of teenagers and adults with FAS/FAE have been studied. In this group, cognitive

impairment was observed. About half of the FAS group had IQ scores falling in the range of mental retardation, FAE patients had somewhat higher average intellectual function, and there was a wide range of IQ scores among both those with FAS and those with possible FAE. Superficially intact verbal skills and academic limitations, including a particular disability in arithmetic, were seen. Attentional deficits and difficulties with judgment, comprehension, and abstraction were frequently found. Maladaptive behaviors and lifelong adjustment problems were common among older individuals with FAS/FAE. There is much to be learned about the lifespan pattern of development of those with alcohol-related birth defects. Researchers are now beginning to carry out neuropsychological assessments of older individuals with FAS, to study in more detail how they function in daily life and to investigate more completely the full spectrum of children, teenagers, and adults with FAS and possible FAE.

FUTURE DIRECTIONS

The effects of prenatal alcohol exposure have been a topic of recent research. Extensive reading lists are available in this volume and in a 1992 article by Carmichael Olson, which include work by the many investigators studying alcohol effects. There is a clear need for continuing study of the lifelong effects of FAS, and the impact of all levels and patterns of drinking during pregnancy on offspring development.

There is great interest in preventing alcohol-related birth defects, identifying individuals with FAS/FAE, and developing successful intervention methods. To accomplish this, a number of steps should be taken. Professionals in the health, educational, and correctional systems need to be made aware of the full spectrum of alcohol-related birth defects. Public awareness is also important, especially among adolescents and women of childbearing age, as well as within groups at risk for heavy alcohol consumption or high incidence of FAS/FAE. Better screening techniques are needed to identify women at high risk for heavy drinking during pregnancy. Instruments for reliable and valid self-report of substance use should be developed, as should sensitive biochemical markers for alcohol consumption. Systems for early identification and high-risk infant monitoring should be established. For alcohol-affected individuals, adequate educational and

vocational assessment, as well as intervention planning, should be developed. There must be development of ongoing funding for services and advocacy for future service needs.

Recently, the Centers for Disease Control made FAS a public health priority, and reducing the incidence of FAS is an objective in the Healthy People 2000 national health promotion strategy. These are important steps in the effort to confront the major public health problem of fetal alcohol syndrome.

BIBLIOGRAPHY

- ABEL, E. L., & SOKOL, R. J. (1987). Incidence of fetal alcohol syndrome and economic impact of FAS-related anomalies. *Drug and Alcohol Dependence*, 19, 51-70.
- CARMICHAEL OLSON, H., BURGESS, D. M., & STREISSGUTH, A. P. (1992a). Fetal Alcohol Syndrome (FAS) and Fetal Alcohol Effects (FAE): A lifespan view, with implications for early intervention. *Zero To Three*, 13(1), 24-29.
- CARMICHAEL OLSON, H., FELDMAN, J. J., STREISSGUTH, A. P., & GONZALEZ, R. D. (1992b, June). *Neuropsychological deficits and life adjustment in adolescents and adults with fetal alcohol syndrome*. Paper presented at the meeting of the Research Society for Alcoholism, San Diego, CA.
- CARMICHAEL OLSON, H., SAMPSON, P. D., BARR, H. M., STREISSGUTH, A. P., & BOOKSTEIN, F. L. (1992c). Prenatal exposure to alcohol and school problems in late childhood: A longitudinal prospective study. *Development and Psychopathology*, 4, 341-359.
- CLARREN, S. K., ASTLEY, S. J., BOWDEN, D. M., LAI, H., MILAM, A. H., RUDEEN, P. K., & SHOEMAKER, W. J. (1990). Neuroanatomic and neurochemical abnormalities in non-human primates exposed to weekly doses of alcohol during gestation. *Alcoholism: Clinical and Experimental Research*, 14(5), 674-683.
- COLES, C. D., BROWN, R. T., SMITH, I. E., PLATZMAN, K. A., ERICKSON, S., & FALEK, A. (1991). Effects of prenatal alcohol exposure at school age, I. Physical and cognitive development. *Neurotoxicology and Teratology*, 13(4), 357-367.
- CONRY, J. (1990). Neuropsychologic deficits in fetal alcohol syndrome and fetal alcohol effects. *Alcoholism: Clinical and Experimental Research*, 14(5), 650-655.
- DAY, N. L., ROBLES, N., RICHARDSON, G., GEVA, D., TAYLOR, P., SCHER, M., STOFFER, D., CORNELIUS, M., & GOLDSCHMIDT, L. (1991). The effects of prenatal alcohol

- use on the growth of children at three years of age. *Alcoholism: Clinical and Experimental Research*, 15(1), 67-71.
- DEPARTMENT OF HEALTH AND HUMAN SERVICES. (1990). *Healthy People 2000: National Health Promotion and Disease Prevention Objectives*. DHHS Publication No. (PHS) 91-50213.
- FEDERAL DRUG ADMINISTRATION. (1981). Surgeon General's advisory on alcohol and pregnancy. *FDA Drug Bulletin*, 11(2), 9-10.
- FRIED, P. A., & WATKINSON, B. (1990). 36- and 48-month neurobehavioral follow-up of children prenatally exposed to marijuana, cigarettes and alcohol. *Developmental and Behavioral Pediatrics*, 11(2), 49-58.
- GIUNTA, C. T., & STREISSGUTH, A. P. (1988). Patients with fetal alcohol syndrome and their caretakers. *Social Casework: The Journal of Contemporary Social Work*, September, 453-459.
- GREENE, T. H., ERNHART, C. B., AGER, J., SOKOL, R. J., MARTIER, S., & BOYD, T. A. (1990). Prenatal exposure to alcohol and cognitive development. *Neurotoxicology and Teratology*, 13(1), 57-68.
- JONES, K. L., & SMITH, D. W. (1973). Recognition of the fetal alcohol syndrome in early infancy. *Lancet*, 2, 999-1001.
- JONES, K. L., SMITH, D. W., ULLELAND, C. N., & STREISSGUTH, A. P. (1973). Pattern of malformation in offspring of chronic alcoholic mothers. *Lancet*, 1, 1267-1271.
- KODITUWAKKU, P. W., HANDMAKER, N. S., CUTLER, S. K., WEATHERSBY, E. K., HANDMAKER, S. D. & AASE, J. M. (1992, June). *Specific impairments of self regulation in FAS/FAE: A pilot study*. Paper presented at the meeting of the Research Society for Alcoholism, San Diego, CA.
- LEMOINE, P., HAROUUSEAU, H., BORTEYRU, J. P., & MENUET, J. C. (1968). Children of alcoholic parents: Abnormalities observed in 127 cases. Available from: *National Clearinghouse for Alcohol Information*, P.O. Box 2345, Rockville, MD, 20852.
- LITTLE, B. B., SNELL, L. M., ROSENFELD, C. R., GILSTRAP, L. C., & GANT, N. F. (1990). Failure to recognize fetal alcohol syndrome in newborn infants. *American Journal of Diseases of Children*, 144, 1142-1146.
- MARTIN, J. C. (1992). *Neurobehavioral teratology: Mental and physiological deficits in offspring as a result of trauma during gestation*. Manuscript submitted for publication.
- NATIONAL INSTITUTE FOR ALCOHOLISM AND ALCOHOL ABUSE. (1990). *Seventh Special Report to the U.S. Congress: Alcohol and Health*. Washington, DC: U.S. Department of Health and Human Services.
- NATIONAL INSTITUTE FOR ALCOHOLISM AND ALCOHOL ABUSE. (1991). Fetal alcohol syndrome. *Alcohol Alert*, 13, PH 297, 1-4.
- NANSON, J. L., & HISCOCK, M. (1990). Attention deficits in children exposed to alcohol prenatally. *Alcoholism: Clinical and Experimental Research*, 14(5), 656-661.
- RANDALL, C. L. (1987). Alcohol as a teratogen: A decade of research in review. *Alcohol and Alcoholism, Suppl. 1*, 125-132.
- RILEY, E. P. (1990). The long-term behavioral effects of prenatal alcohol exposure in rats. *Alcoholism: Clinical and Experimental Research*, 14(5), 670-673.
- ROSETT, H. L. (1980). A clinical perspective of the fetal alcohol syndrome. *Alcoholism: Clinical and Experimental Research*, 4(2), 119-122.
- RUSSELL, M. (1991). Clinical implications of recent research on the fetal alcohol syndrome. *Bulletin of the New York Academy of Medicine*, 67(3), 207-222.
- RUSSELL, M., CZARNECKI, D. M., COWAN, R., MCPHERSON, E., & MUDAR, P. J. (1991). Measures of maternal alcohol use as predictors of development in early childhood. *Alcoholism: Clinical and Experimental Research*, 15, 991-1000.
- SCHENKER, S., BECKER, H. C., RANDALL, C. L., PHILLIPS, D. K., BASKIN, G. S., & HENDERSON, G. I. (1990). Fetal alcohol syndrome: Current status of pathogenesis. *Alcoholism: Clinical and Experimental Research*, 14(5), 635-647.
- SOKOL, R. J., & CLARREN, S. K. (1989). Guidelines for use of terminology describing the impact of prenatal alcohol on the offspring. *Alcoholism: Clinical and Experimental Research*, 13(4), 597-598.
- STREISSGUTH, A. P. (1992). Fetal alcohol syndrome and fetal alcohol effects: A clinical perspective of later developmental consequences. In I. S. Zagon & T. A. Slotkin (Eds.), *Maternal substance abuse and the developing nervous system*. San Diego: Academic Press.
- STREISSGUTH, A. P., AASE, J. M., CLARREN, S. K., RANDELS, S. P., LADUE, R. A., & SMITH, D. F. (1991). Fetal alcohol syndrome in adolescents and adults. *The Journal of the American Medical Association*, 265(15), 1961-1967.
- STREISSGUTH, A. P., BARR, H. M., & SAMPSON, P. D. (1990). Moderate prenatal alcohol exposure: Effects on child IQ and learning problems at age 7½ years. *Alcoholism: Clinical and Experimental Research*, 14(5), 662-669.
- STREISSGUTH, A. P., & CLARREN, S. K. Personal communication. December 11, 1992.
- STREISSGUTH, A. P., & LADUE, R. A. (1987). Teratogenic causes of developmental disabilities. In S. R. Schroeder

(Ed.), Toxic substances and mental retardation, *AAMD Monographs*, 8, 1–32.

ULLELAND, C. N. (1972). The offspring of alcoholic mothers. *Annals of the New York Academy of Sciences*, 197, 167–169.

WEST, J. R. (1986). *Alcohol and Brain Development*. New York: Oxford University Press.

WEST, J. R., GOODLETT, C. R., BONTHIUS, D. J., & PIERCE, D. R. (1990). New approaches to research on the long-term consequences of prenatal exposure to alcohol. *Alcoholism: Clinical and Experimental Research*, 14(5), 684–689.

HEATHER CARMICHAEL OLSON

FLUID AND CRYSTALLIZED INTELLIGENCE, THEORY OF Evidence accumulated over the course of the twentieth century has made it clear that the phenomenon of human intelligence is multidimensional. A system for explaining this evidence is the theory of fluid (*Gf*) and crystallized (*Gc*) intelligence, known simply as *Gf-Gc theory*. This theory has changed with the accumulation of evidence. At first it was a theory of two intelligences; today, it would be better labeled the theory of many intelligences (see MULTIPLE INTELLIGENCES). But the term *Gf-Gc theory* has become ensconced in language and will not be easily phased out.

The evidence from which the theory derives is of five principal forms:

1. evidence of individual differences, called *structural evidence*
2. evidence of change from infancy to old age, called *developmental evidence*
3. evidence of relationships to indicators of physiological and neurological functioning, called *neurocognitive evidence*
4. evidence of predictions of school performance and occupational levels, called *evidence of achievement*
5. evidence of relationship among persons related biologically in different degrees, called *heritability* or *behavioral-genetic evidence*

These five forms of evidence indicate that a single scientific concept does not represent the phenomena: what is referred to as *intelligence* is a mélange of many different cognitive capabilities.

Figure 1 provides a schematic indication of the descriptive concepts of the theory. Briefly described, the concepts are (from top to bottom in the figure):

Fluid Reasoning (Gf), measured in tasks requiring inductive, deductive, conjunctive, and disjunctive reasoning to arrive at understanding of relations among stimuli, to comprehend implications, and to draw inferences.

Acculturation Knowledge (Gc), measured in tasks indicating breadth and depth of the knowledge of concepts and forms of reasoning that have been developed by humans over the course of many centuries and passed on from one generation to the next. *Gc* can be thought of as the intelligence of the culture that is incorporated by individuals through a process of acculturation.

Visual Processing (Gv), measured in tasks involving visual closure and constancy, and fluency in “imaging” how objects can appear in space as they are rotated and flip-flopped in various ways.

Auditory Processing (Ga), measured in tasks that involve perceiving sound patterns under distraction or distortion, maintaining awareness of order and rhythm among sounds, and comprehending elements of groups of sounds, such as chords, and the relations among such groups.

Processing Speed (Gs), involved in almost all intellectual tasks (Hertzog, 1989) and regarded by some (e.g., Eysenck, 1982) as the central feature of intelligence, measured most purely in rapid scanning in intellectually simple tasks (in which almost all people would get the right answer if the task were not highly speeded).

Correct Decision Speed (CDS), often regarded as central to intelligence, measured in quickness in providing answers in tasks that require the solving of problems.

Short-term Apprehension-Retention (SAR), also called *short-term memory (GSM)*, sometimes re-

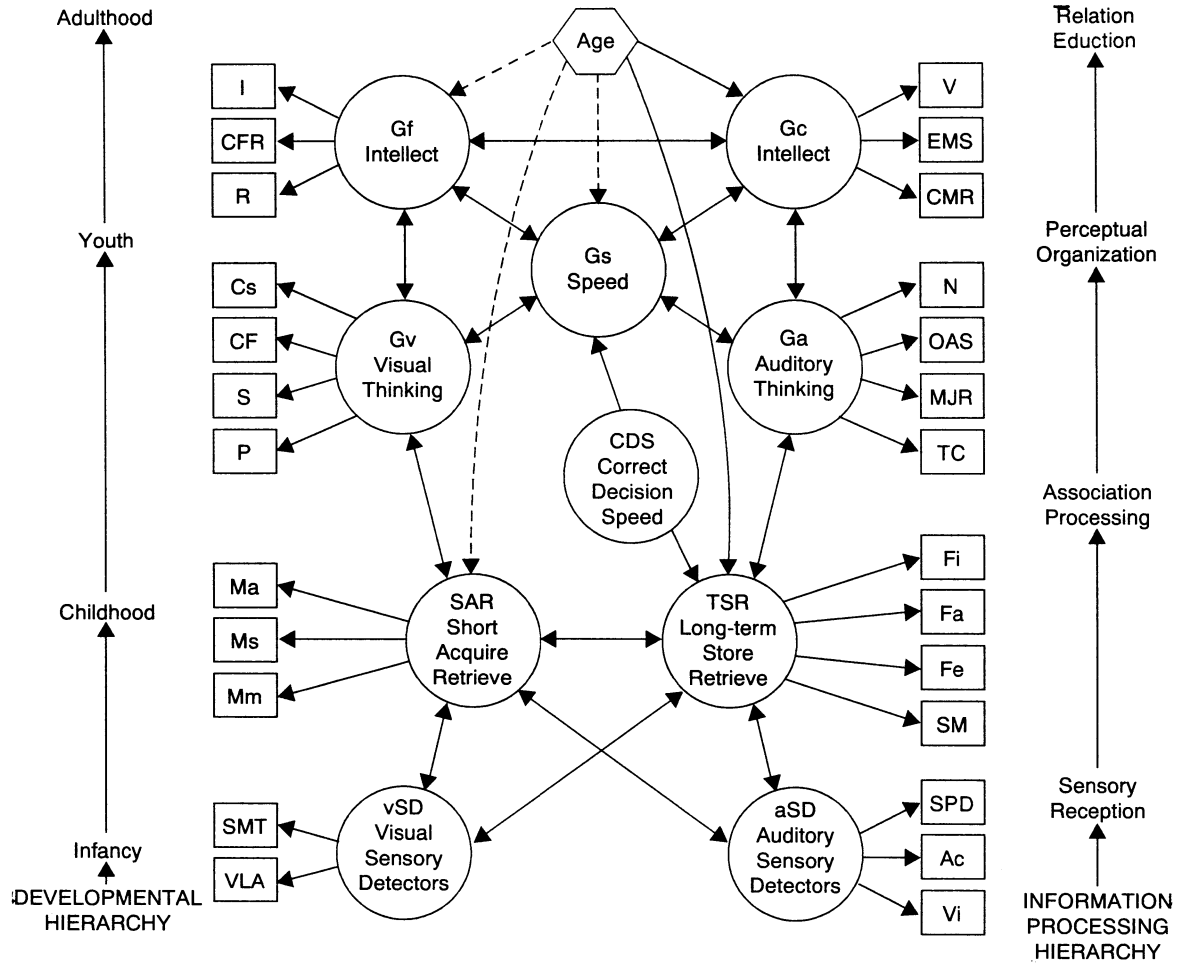


Figure 1
Intellectual ability organization at a broad level

I	Inductive Reasoning	Tc	Temporal Tracking
CFR	Cognition of Figural Relations	Ga	Broad Auditory Comprehension
R	Quantitative Reasoning	Ma	Associational Immediate Memory
Gf	Fluid Education	Ms	Span Immediate Memory
V	Verbal Comprehension	Mm	Meaning-Paired Associates Immediate Memory
EMS	Evaluation of Semantic Systems	SAR	Short-Term Apprehension and Retrieval
CMR	Cognition of Semantic Relations	Fi	Ideational Fluency
Gc	Acculturated Knowledge	Fa	Associational Fluency
Cs	Speed of Closure	Fe	Expressional Fluency
Cf	Flexibility of Closure	SM	Semantic Memory Over Minutes
S	Spatial Orientation	SMT	Sperling Matrix Awareness
P	Perceptual Speed	VLA	Visual Location Address
Gv	Broad Visual Comprehension	vSD	Visual Sensory Detection
Gs	Broad Speediness	SPD	Speech Perception Under Distraction
CDS	Correct Decision Speediness	Ac	Auditory Acuity
DAS	Discrimination Among Sound Patterns	Va	Auditory Valence Recall
MJR	Maintaining and Judging Rhythms	aSD	Auditory Sensory Detection

garded as the central feature of intelligence, measured in a variety of tasks that indicate awareness of, and ability to recall, elements of immediate stimulation—i.e., events in the passing parade of the past minute or so. Immediate memory is rather like a register into which awarenesses come and are passed on either to storage or to oblivion.

Fluency of Retrieval from Long-Term Storage (TSR), also called *long-term memory (Glr)*, and sometimes regarded as intelligence, measured in tasks that indicate consolidation for storage and mainly require retrieval, through association, of information that was stored minutes, hours, weeks, and years before.

Visual Sensory Detection (vSD), measured in tasks in that indicate fleeting awareness (over a few milliseconds) of a large amount of visual information—curvature, symmetry, parallelism, edges, wedges, blocks, cones (Biederman, 1987; Sperling, 1960).

Auditory Sensory Detection (aSD), sometimes called *echoic memory*, measured in tasks requiring awareness for very short periods of time (milliseconds) of relatively large amounts of auditory information.

The concepts of Figure 1 correspond to second-order abilities indicated by structural evidence. The little boxes represent primary mental abilities which each second-order ability comprises. For example, *Gc* encompasses primary abilities of Verbal Comprehension (*V*), Evaluation of Semantic Systems (*EMS*), Comprehension of Semantic Relations (*CMR*), Deductive Reasoning, and Numerical Facility, as well as various knowledges measured in achievement batteries. Other *Gf-Gc* factors similarly involve several different elementary abilities.

A separate concept, also, is *quantitative knowledge (Gq)*, an understanding of the concepts and skills of mathematics. *Gq* is a form of crystallized knowledge, but individual differences in quantitative abilities are so notably different from individual differences in other broad bands of knowledge that it is necessary to

recognize them as involving processes of acquisition and retention that are different from those for *Gc*. A circle representing *Gq* would appear near the top in Figure 1.

The precise number of distinguishable cognitive capabilities evinced by humans has not been established. Indeed, such a number probably never will be determined because humans manifest myriad of such capabilities; this number is expanding and contracting as evolution and human history progress; these capabilities represent many, many different functions; and this information can be organized in a great many ways, none of which is fully exhaustive. The situation is somewhat analogous to the impossibility of specifying all the distinct chemical substances that occur in the universe.

In chemistry there is a system of elements in terms of which all substances (known and yet-to-be-determined) can be described. Thus, although the number of all distinct chemical substances cannot be determined, the nature of all such substances can be understood in terms of a finite number of elements. In psychology no such system of elements has been discovered. Nevertheless, the evidence is clear that among a myriad of human cognitive capabilities there are common features. This suggests that a finite system of elements—capacities and processes—does bind cognitive capacities and in terms of which all of the myriad of human abilities can be understood—even as we have yet to find such a fully inclusive system.

Major considerations for such a system have been put forward. Some of these are discussed here.

STRUCTURAL EVIDENCE

Structural evidence derives from hundreds of studies of the relationships among measures of individual differences in human cognitive capabilities. In these studies (and in the hundreds of studies on which they are based), literally thousands of tests, experimental procedures, and paradigms were developed to measure what appeared to be quite distinct and separate abilities, capacities, and cognitive processes. Yet a surprising finding of much research is that among this huge diversity of different measures there is a large amount of redundancy: tests that would seem to measure different abilities are found to measure the same abilities

in common. The tests have different names, but they do not measure different abilities.

It's as if you were to line people up according to height and find that you had also lined them up according to amount of money in the bank, years of education, distance from home, and a whole host of other things. You wouldn't expect such an outcome, and if you found it, you would be surprised. Similarly, psychologists constructed measures of what they thought were different abilities, but found that the tests ordered individuals in much the same way. They found that a large proportion of variability produced by a very large number of measures of human capabilities could be reliably described in terms of relatively small number of common factors (see FACTOR ANALYSIS).

Common factors among abilities have been found at four levels. The number of factors now established at what is known as the primary mental ability level is somewhat greater than forty (see PRIMARY MENTAL ABILITIES). These abilities account for redundancy in measures of many hundreds of tests. But these forty-plus abilities are not entirely independent: they, too, order individuals in somewhat the same way. When the redundancy among these primary abilities is analyzed, evidence of approximately ten factors is found at what is known as the second order. And that's not all. These ten factors are also not entirely independent: when their redundancy is analyzed, two common factors are found. And that's still not all. These factors are positively correlated, which indicates one common factor.

Thus it is that individual differences in a huge number of human cognitive capabilities can be described to a large extent in terms of a system of approximately forty primary abilities, which in turn can be described in terms of ten broad abilities, which themselves indicate two very broad abilities, which indicate one very, very broad ability. As surprising and important as this finding is, it is only a glimmering of the organization among human capacities. The evidence of developmental, neurocognitive, achievement, and behavioral-genetic research also indicates common features among many functions associated with cognitive capabilities. These common features are only partially understood at this point in history, but the findings from many perspectives converge to indicate that what

is seen in individual differences is seen also in development, life achievements, neurological functioning, and biogenetic relationships.

THE COMPROMISE OF *Gf-Gc* THEORY IN RELATION TO IQ

Gf-Gc theory focuses on the second-order abilities revealed in structural analyses. This is a compromise. Ideally, theory should deal with the developmental, life achievement, neurological, and genetic relationships of the forty-plus primary abilities, but this is an extremely difficult task, and the empirical evidence on which to build such a theory is scanty. The research really needed to produce such evidence is hugely complex and expensive. It has yet to be done. On the other hand, it is clear that a theory of just one factor—general intelligence—is not adequate to explain the phenomena and the evidence that have been adduced. The compromise of *Gf-Gc* theory is between these extremes: it is an account of the common factors among the primary mental abilities, and of the interrelationships among these broad forms of intelligence.

The Wechsler and Stanford-Binet tests, as well as most other commercial IQ and neuropsychological (also called neurocognitive) batteries, measure some of the second-order abilities of Figure 1 (see GF-GC MEASURES OF INTELLIGENCE). They measure very few, indeed, of the forty-plus primary abilities of which the second-order abilities are comprised. It is thus evident that what is measured in IQ tests is far from all there is to intelligence: IQ tests measure very limited samples of the known, measurable abilities of intelligence.

Each of several of the second-order abilities of Figure 1 has, on its own, been regarded as equal to, or central to, general intelligence (IQ, or *g*). Indeed, *Gf* and *Gc* have been discussed as if they are equivalent—as if each indicates the same IQ or *g* (Jensen, 1984). Short-term apprehension and retrieval and quick thinking, too, have been regarded as the central feature of intelligence. But none of these capabilities well represents all the others or is the essence of all the phenomena referred to with the word *intelligence*. They are better regarded as distinct intelligences, which is not to say that they are unrelated. A person functions as a whole, not as merely a collection of unrelated parts.

CONSTRUCT VALIDITY

Structural evidence indicates the distinctiveness of the different capabilities. It indicates that the abilities are construct independent, meaning that each capability is reliably distinct and separate from the other capabilities (and any combination of these). Construct independence is important because it means that what can be established as true for one construct (capability) will not be redundantly the same as what can be established as true for other constructs.

But the distinctiveness of the separate concepts of Figure 1 is demonstrated, also, and more compellingly, by evidence of distinct construct validities (see VALIDITY). This is evidence of how different capabilities are different in their relationships with other variables—how they develop differently and relate differently to variables of physiological structure and function, and those of education and genetics. Evidence of construct

validity is evidence showing that different capabilities stem from different sets of determinants—including different sets of genes—and are affected in different ways by influences associated with injuries, child rearing, education, and the variety of practices that make up different lifestyles.

The results summarized in Figure 2 illustrate evidence of construct validity obtained from studies of development. The curves of the figure are for means (averages) at different ages from adolescence to old age. They show that there is (on the average and across many individuals) monotonic decrease in some intellectual abilities and monotonic increase in other such abilities. The averages on which this particular figure is based are for different individuals at each age (i.e., are cross-sectional differences; see AGING AND INTELLIGENCE). But the same kinds of results are seen also in averages based on repeated measures of the same people (i.e., in longitudinal data). Summarized over

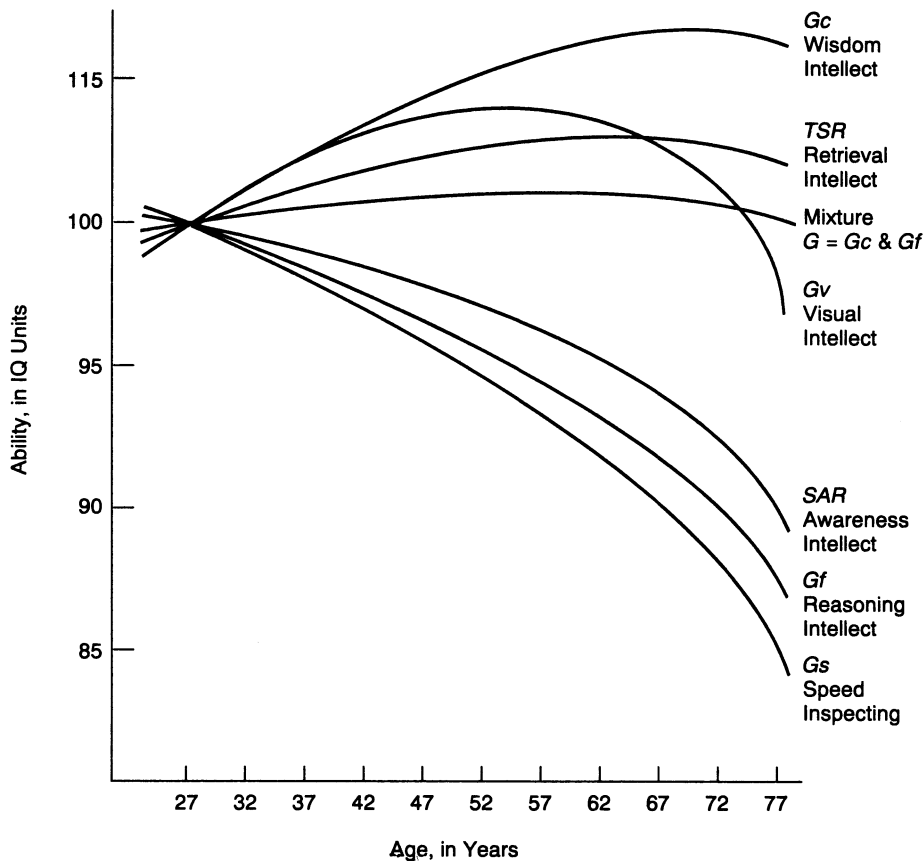


Figure 2
Adulthood age differences in different dimensions of intellect: Schematic summary of eight studies

many studies, the finding is that the averages for G_f , G_s , SAR , and G_v decrease with increasing age in adulthood, while in the same samples of people the averages for G_c and TSR increase. (In a period of very old age, the averages for these latter abilities may decrease also, see Schaie and Baltes, 1977).

Different IQ tests are made up of different mixtures of the G_f - G_c abilities. If such a mixture happens to comprise primarily abilities for which averages decrease with age in adulthood, then the finding of a study of age differences in IQ is that IQ declines with aging. On the other hand, if the mixture comprises primarily abilities for which the averages increase, then the finding is that IQ increases with aging. If the two kinds of abilities are about equally weighted in an IQ test (represented by g in Figure 2), the finding is that IQ neither increases nor decreases with age in adulthood.

Such findings, interpreted as "conflicting," illustrate why IQ (general intelligence, g) is not a good scientific construct. There is no real conflict in the results obtained with different IQ tests: it is misleading to interpret the findings as conflicting. Distinct developmental processes are at work in each of the different cognitive capabilities that are arbitrarily mixed (in different proportions) in different measures of IQ. These processes are confounded—not identified—in studies of IQ. Such confounding hides not only developmental differences but also differences in relationships with neurological, achievement, and genetic variables.

DEVELOPMENTAL EVIDENCE IN RELATION TO FUNCTION

Returning to Figure 1, notice that from bottom to top in the figure are two important hierarchies—one of development, and one of function (information processing). With regard to function, Figure 1 illustrates that sensory detection capabilities (represented at the bottom of the figure) feed into and support association processing at the next level, which processes feed into and support the organizational functions represented by G_v , G_a , and G_s , which in turn are necessary for exercise of the relational thinking functions (eduction of relations and correlates) of G_f and G_c .

Figure 1 also represents the idea that development parallels the functional hierarchy. The first responses to become evident in the infant indicate sensory awareness, the lowest level of function. Repetitious activities then develop into short-term memories, some of which are consolidated in long-term storage. Through massive thinking and rethinking over years of development, memories are organized and reorganized again and again in visual and auditory concepts that are identified in G_v , G_a , G_f , G_c , and the retrieval functions of TSR . The development of cognitive capabilities thus parallels the functional organization of these capabilities. Also, the evidence suggests that in cognitive development and function, as elsewhere in biological development and function, ontogeny recapitulates phylogeny. The following is a brief outline of this development.

There is very little in the infant's behavior that relates to what is later referred to as *intelligence*. Yet even before birth, the infant can sense the movement of sound, and shortly after birth, it demonstrates that it can see and hear. These demonstrations indicate sensory awareness. The infant's behavior in the weeks following birth indicates increasingly that it responds to events occurring a few seconds before the response: this indicates short-term apprehension and retrieval (SAR). Also increasingly, the infant indicates that it recognizes images, and patterns of sounds, touches, and tastes similar to what it has experienced minutes, hours, and weeks before. This indicates awareness of the past (storage, as in G_c) and retrieval from storage (TSR). It becomes clear in the first year of life that these elementary apprehensions and memories are organized with respect to both visual and auditory stimulation. Indicating individual differences, infants respond to music—indicating G_a —in a different way than they respond to a visual stimulation—indicating G_v .

Two important forms of cognitive processing can be seen to emerge in the first year of the infant's life: one is consolidation and cognitive storage; the other is directed exploration. In *PIAGETIAN THEORY OF DEVELOPMENT* these processes are described under the heading of accommodation and assimilation. They appear to be the early indications of G_c and G_f , respectively.

Consolidation is learning that is retained, stored,

and available through retrieval over long periods of time. It occurs when there is understanding of that which is learned. It can occur with practice. An early form of practice is seen in what are called circular reactivities, which are prominent in the infant's behavior in the first year of life. Circular reactivities are sequences of behavior repeated over and over. It is as if the infant is practicing—and this behavior appears to be rewarding. The overt circular reactivity is a manifest indication of subjective (thinking) circular reactivity. Just as practice is a basis for consolidation in learning, so circular reactivity is a basis for consolidation. This is indicated by recognition of stimulus patterns similar to patterns presented previously. Such recognition indicates subjective concept awareness: in distinguishing the familiar from the unfamiliar, the infant displays concept awareness. Concepts are the elements of crystallized knowledge (G_c). Awareness of concepts is seen with increasing frequency as the child nears the end of the first year of life. As development proceeds in the second year, concepts increasingly are linked to verbal tags—spoken words. More and more with advancing age in childhood, knowledge of words becomes a measure of G_c .

Near the end of the first year of life, infants display stranger anxiety and in other ways demonstrate that they rather sharply distinguish the familiar from the unfamiliar. They also display surprise, which is similar to stranger anxiety, except that it is rewarding rather than aversive. The infant seeks more of it; it seeks novelty. This seeking seems to be driven by the rewards of surprise. Children explore their surroundings with expectations built up in memory. They display surprise when they encounter departures from expectations.

Novelty seeking, or directed exploration, is thus rather like problem solving: it is exploring with “hope” of being rewarded. It appears to be one of the early indications of fluid reasoning: it is predictive of IQ measured in the preschool period.

Distinctions between different cognitive capabilities become more and more prominent as development proceeds through the third year of life. By age 4 it is clear that early forms of the cognitive capabilities of Figure 1 can be distinguished (e.g., Horn, 1986, Table 2.5).

Among commercial tests, Woodcock-Johnson is the only one that provides childhood measures for all the second-order abilities, but good measures of G_f , G_j , SAR , and TSR can be obtained with the McCarthy battery. The Kaufman batteries provide good measures of G_c , G_v , and SAR —in the KABC—and G_f , as well as G_c , in the adult test. (This encyclopedia contains articles describing each of these tests. See, particularly, DEVELOPMENT, COGNITIVE.)

EVIDENCE OF ACHIEVEMENTS

G_f , G_c , and other second-order abilities predict grades in school, ratings of performance in many kinds of training, and, in general, many kinds of achievements in which intellectual abilities are required. The predictive correlation in these situations is typically about .50 for G_c , about .40 for G_f , and .30 or smaller for other abilities.

Measures of outcomes (grades, ratings, on-the-job tests) are usually broad and complex: performances that involve several different abilities used in a variety of ways. Because IQ tests involve several different abilities, they often predict achievement criteria better than relatively pure measures of G_c or G_f or other second-order abilities.

NEUROCOGNITIVE EVIDENCE

Abilities increase with learning, practice, and use. Without use, they decline. They decline also with loss of neurological base.

Over the ages of childhood, the averages for all cognitive abilities increase. As development continues through adulthood, the averages for G_c and TSR continue to increase, but the averages for G_f , SAR , and G_s decrease. The latter are said to be vulnerable abilities. Not only are these the abilities that decline first and most with age in adulthood, they are also most irreversibly affected by—or are most vulnerable to—injuries to the central nervous system (CNS). The evidence is less clear in suggesting that the sensory detectors G_a and G_v also decline, and there is no good evidence that CDS either does or does not decline.

The findings of decline associated with aging and with known brain damage suggest that aging involves brain damage. It seems that with aging there is an ac-

cumulation of CNS injuries over the life course. Many things can cause such injury—high fever, inhalation of carbon monoxide, inebriation, the “highs” and “lows” produced by taking different drugs (prescription as well as illicit), anesthetics, hypertension, and blows to the head. Each such injury might be small, but CNS damage is irreversible, so even small effects can accumulate to produce the net effect of one large injury such as can be seen in cases of known brain damage. Thus the accumulation of small injuries can produce the aging declines seen in *Gf* and the other vulnerable abilities.

In contrast to the vulnerable abilities, *Gc* and *TSR* are maintained abilities: they do not decline with age in adulthood. Also, although these abilities are depressed immediately following brain damage (such as that produced by stroke), they “spring back” nearly to pre-injury level in the weeks of the recovery period.

These findings lead to a hypothesis that *Gf* and *TSR* are maintained through processes of overdetermination and equipotentiality. These abilities are practiced over and over again under many, many different conditions. This ties many stimulus cues to cognitive associations and to synaptic links between neurons. Such cognitive associations and neural strands become interwoven in complex networks in which any of several strands holds the network together. This is overdetermination. In an overdetermined network, loss of some strands through brain damage does not destroy the ability of cues to activate the network because many strands remain: the entire pattern can be activated through any of many stimulus cues. Thus, abilities are maintained despite some loss of the neural base.

At a neurological level, abilities are maintained until injury becomes massive enough, through accumulation or with a single injury (such as stroke), to destroy such a large segment of the network that the few remaining strands are not likely to be activated by any of the few cognitive cues that can activate the system.

Unless the damage is very extensive indeed, however, overdetermination is not an all-or-nothing condition: It is a matter of degree. As long as a few neural strands remain, there are a few cues that can activate a goodly portion of the remaining network, with the result that the person may at times seem to have abilities that at other times appear to be lost. This can be seen even in extreme cases of Alzheimer’s disease.

At the level of practice and use, overdetermination is lost with inactivity, which is to say when abilities are not practiced. Even if there is no neural loss, and even with good consolidation in learning and much overdetermination with use, an ability that is not used will decline. The curve indicating the rate of such loss is not as steep as is seen in the familiar learning curve of recently learned—usually not well-consolidated—skills, but the shape is similar. Some of the decline in abilities seen at any age is loss resulting from lack of use. Such decline can be quite noticeable, as when active people “drop out” of a culture (fail to use a language they formerly knew, for example). Some of the decline seen in old age may stem from such inactivity.

HERITABILITY EVIDENCE

Results from behavioral-genetic research indicate that *Gf* and *Gc* have distinct heritabilities (see HERITABILITY). Early studies were directed at establishing the extent to which IQ (general intelligence) was inherited. The results were confusing. The heritability estimates of different studies ranged from 30 to 90 percent. Some of the confusion resulted from the same factors that produced confusion in interpreting results from studies of aging: Different collections of abilities were incorporated in the IQ measures of different studies. It was thought that the *Gf-Gc* distinction would reduce this confusion in much the same way as it did for results on aging. Although this remains a logically compelling possibility, the evidence thus far accumulated has not supported it.

R. B. CATTELL (1941; 1957) put forth the basic hypotheses. He suggested that *Gf* reflects primarily genetic influences, and *Gc* reflects mainly environmental influences. Cattell’s thought was that the genetic potential of *Gf* is “invested” in *Gc* over the course of development, but *Gc* was also produced through environmental conditions for which there are individual differences. *Gc* would be perfectly correlated with *Gf* were it not for differences in environmental factors through which intelligence is developed, but it becomes independent from *Gf* because individual differences in environmental influences occur and accumulate throughout childhood and subsequently (Horn & Cattell, 1966). Thus, if an IQ test comprised

mainly G_c measures, it would have low heritability—say, 30 percent. But if an IQ test was made up mainly of G_f tests, it would have high heritability (e.g., 80 percent).

It follows from this reasoning that in the earliest period of life there will be virtually no distinction between G_f and G_c —because there will have been few individual differences in environmental influences and little time for such influences to operate and accumulate—but that as development proceeds beyond the earliest years, the distinction between G_f and G_c will become clearer and clearer. It follows, too, that for equally reliable measures and representative sampling of subjects, the heritability of G_f should be larger than the heritability of G_c .

Most of the evidence bearing on these questions does not support either of the main arguments of Cattell's hypotheses—that G_f will have higher heritability than G_c , and that G_f and G_c will not be distinguishable early in life. In eleven separate studies based on analyses of monozygotic and dizygotic twins, the heritability estimates for G_f abilities were not systematically larger than the heritability estimates for G_c abilities. The results are consistent with a hypothesis that G_f and G_c stem from different genetic determinants, the effects of which can be seen early in development.

Among 4-year-old children, G_f and G_c factors are quite distinguishable (as are some other second-order abilities), and the correlation between the two is no larger than has been found in adulthood studies. This evidence does not support the hypothesis that G_f and G_c are indistinguishable at an early age and become more and more separable as development proceeds.

OTHER CONSIDERATIONS

One might wonder why there are no concepts in G_f - G_c theory to represent organization among other sensory functions, such as those of taste and smell and kinesthesia. A complete theory would have an account of these capabilities, for it is evident that just as one thinks in terms of sights and sounds, so one also thinks in terms of tastes and smells and "feels." But little is known scientifically about these capabilities, and particularly about how they relate to other intellectual functions. It is therefore premature to say much more than that it is important to recognize that human in-

telligence involves more than what we experience visually and auditorially.

Most of the abilities identified as cognitive capabilities are positively correlated. This has led to the thought that a single process—called *general intelligence*, often symbolized as either IQ or g —underlies human cognitive capability. If there were such a single process, G_f would be a salient indicator of that feature. But as reviewed in this article, the evidence does not point to such a single feature. G_f does not account for—does not predict—the major proportion of individual differences in all intellectual abilities. It is only one part of the broad conglomerate of all abilities that researchers have identified as indicating intelligence.

BIBLIOGRAPHY

- BIEDERMAN, I. (1987). Recognition-by-components: a theory of human image understanding. *Psychological Review*, 94, 115–147.
- CATTELL, R. B. (1941). Some theoretical issues in adult intelligence testing. *Psychological Bulletin*, 38, 592.
- CATTELL, R. B. (1957). *Personality and motivation structure and measurement*. New York: World Book.
- EYSENCK, H. J. (ED.). (1982). *A model for intelligence*. New York: Springer-Verlag.
- HERTZOG, C. (1989). Influences of cognitive slowing on age differences. *Developmental Psychology*, 25, 636–651.
- HORN, J. L. (1986). Intellectual ability concepts. In R. L. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence* (Vol. 3, pp. 35–77). Hillsdale, NJ: Erlbaum.
- HORN, J. L., & CATTELL, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, 57, 253–270.
- JENSEN, A. R. (1984). Test validity: g versus the specificity doctrine. *Journal of Social and Biological Structures*, 7, 93–118.
- SCHAIK, K. W., & BALTES, P. B. (1977). Some faith helps see the forest: A final comment on the Horn and Donaldson myth of the Baltes-Schaeie position on adult intelligence. *American Psychologist*, 32, 1118–1120.
- SPERLING, G. (1960). The information available in brief visual presentations. *Psychological Monographs*, 74, 498.

JOHN L. HORN

FLUID AND CRYSTALLIZED THEORY OF INTELLIGENCE, MEASURES OF

Since 1941, the *Gf-Gc* theory of fluid and crystallized intelligence has emerged as a major empirically based organization of evidence indicating that intelligence is a composite of several broad abilities that work in concert to produce cognitive performance. At least eight such broad abilities have been identified and described in the research of Raymond B. CATTELL, John Horn, John B. CARROLL, and others. They have been labeled as follows: short-term memory (*Gsm*), comprehension-knowledge (*Gc*), quantitative ability (*Gq*), visual processing (*Gv*), auditory processing (*Ga*), associative storage-retrieval (*Glr*), novel reasoning (*Gf*), and processing speed (*Gs*).

The research whereby these abilities have been identified was based on test batteries designed to investigate hypotheses about the nature of human cognitive capabilities and to describe the organization of human abilities. Rarely have these studies included the subtests, as such, of tests put together and published to measure intelligence for use in counseling, diagnosis, evaluation, and guidance. Typically, the latter are individually administered tests, whereas most of the tests used in research are administered in groups. But some of the tests of intelligence have been used in basic research, and some research has been directed at identifying the abilities measured in intelligence-test batteries. The results from this work indicate that the eight abilities derived from basic research are measured more or less well by subtests included in the major intelligence batteries currently used in the United States and countries in which English is the dominant language, as well as in other countries. The determination of which subtests measure which *Gf-Gc* abilities, and how well, may be addressed through use of common FACTOR ANALYSIS.

FACTOR-ANALYTIC STUDIES

Two requirements must be met in a factor-analytic study that indicates which tests in a battery indicate which abilities. First, a breadth of cognitive abilities must be represented in the study. Second, there must be a sufficient number (generally three or more) of measures (markers) for each of the factors. These re-

quirements have not been met in most of several dozen factor-analytic studies in the literature that purport to describe the factorial structure of published batteries of tests of intelligence. The abilities reliably measured in these studies have been restricted to those subtests included within the published battery itself. As a result, many of the major abilities required to perform the various subtests have not been differentiated, or even detected, in the majority of studies.

The information in this article is based on results from fifteen factor-analytic studies on nine data sets including sixty-eight subtests from four individually administered intelligence batteries (Woodcock, 1990, 1993). The comprehensiveness of the data meets the requirements described above for an analysis of subtest-to-factor relations.

EXAMPLE MEASURES OF *Gf-Gc* FACTORS

The selected examples in Table 1 illustrate the kinds of measures (of the four intelligence batteries) associated with each of the broad *Gf-Gc* abilities. The subtests are ordered by size of factor-loading relationship between test and ability. The relative factor loading provides an approximation of that subtest's strength, among all subtests in the four batteries, in measuring that factor. Not all subtests from the four batteries are listed, just those for which the factor loadings are highest and those which are illustrative (see Woodcock, 1990, pp. 242–243, for a complete listing). The four batteries are the *Kaufman Assessment Battery for Children* (K-ABC), the *Stanford-Binet Intelligence Scale, 4th Edition* (SB IV), the *Wechsler Intelligence Scale for Children, Revised* and the *Wechsler Adult Intelligence Scale, Revised* (WECH), and the *Woodcock-Johnson Tests of Cognitive Ability, Revised* (WJ-R).

Each of the four intelligence batteries provides some system of factor scores, all with different labels. Table 2 reports the congruence of these factor scores with the eight *Gf-Gc* abilities. Certain factor scores provided by some of the batteries are not included in Table 2, as they are not verified in the results of factor studies meeting the requirements of breadth and depth of marker variables (see Woodcock, 1990, pp. 248–250, for further details).

TABLE 1
Example measures of *Gf-Gc* factors from four contemporary intelligence batteries

<i>Factor/ Battery</i>	<i>Subtest</i>	<i>Relative Factor Loading</i>	<i>Description</i>
SHORT-TERM MEMORY (<i>Gsm</i>)			THE ABILITY TO HOLD INFORMATION IN IMMEDIATE AWARENESS AND THEN USE IT WITHIN A FEW SECONDS. MOST <i>Gsm</i> TESTS MEASURE THE SPAN OF INFORMATION THAT CAN BE HELD IN IMMEDIATE AWARENESS.
SB IV	Memory for digits	.78	Part 1 requires the subject to repeat a series of random digits in the same order as presented. Part 2 requires the subject to repeat a series of random digits backward.
WJ-R	Memory for words	.78	Measures the ability to repeat lists of unrelated words in the correct sequence.
K-ABC	Number recall	.72	Measures the ability to repeat a series of random digits in the same order as presented.
WECH	Digit span	.69	Similar to the SB IV memory for digits.
COMPREHENSION-KNOWLEDGE (<i>Gc</i>)			A PERSON'S BREADTH AND DEPTH OF KNOWLEDGE. IT INCLUDES THE ABILITY TO COMMUNICATE (ESPECIALLY VERBALLY) AND THE ABILITY TO REASON USING PREVIOUSLY LEARNED PROCEDURES.
SB IV	Vocabulary	.81	Items 1-14 are picture vocabulary; items 15-46 are oral vocabulary requiring a synonym or definition.
WECH	Vocabulary	.81	Measures the ability to provide definitions for words.
WJ-R	Picture vocabulary	.75	Measures the ability to name familiar and unfamiliar pictured objects.
K-ABC	Riddles	.70	Measures the ability to name a concrete or abstract concept when given a list of its characteristics.
QUANTITATIVE ABILITY (<i>Gq</i>)			THE ABILITY TO MANIPULATE NUMERIC SYMBOLS AND TO REASON PROCEDURALLY WITH QUANTITATIVE INFORMATION AND RELATIONSHIPS.
WJ-R	Calculation	.79	Measures the subject's skill in performing mathematical calculations.
SB IV	Equation building	.78	Measures the ability to organize a set of number operations into a mathematical sentence or equation.
WECH	Arithmetic	.75	Measures the subject's skill in analyzing and solving practical problems in mathematics that are presented orally.
K-ABC	Arithmetic	.62	Measures knowledge of numbers and mathematical concepts, counting and computation skills.
VISUAL PROCESSING (<i>Gv</i>)			THE ABILITY TO ANALYZE AND SYNTHESIZE VISUAL STIMULI.
K-ABC	Triangles	.67	Measures the ability to assemble several identical triangles into a pattern that matches a model.
WECH	Object assembly	.62	Measures the ability to assemble pieces forming a picture of an object that has not been identified.
	Block design	.58	Measures the ability to reproduce a printed design using a set of varicolored cubes.

FLUID AND CRYSTALLIZED THEORY OF INTELLIGENCE, MEASURES OF

<i>Factor/ Battery</i>	<i>Subtest</i>	<i>Relative Factor Loading</i>	<i>Description</i>
SB IV	Pattern analysis	.57	Measures the ability to reproduce a pattern presented by the examiner. Some items utilize a formboard; others utilize cubes that are assembled to duplicate the examiner's model or a printed design.
WJ-R	Visual closure	.47	Measures the ability to name a drawing or picture of a simple object that is altered in one of several ways.
AUDITORY PROCESSING (<i>Ga</i>)			THE ABILITY TO ANALYZE AND SYNTHESIZE AUDITORY STIMULI.
WJ-R	Sound blending	.69	Measures the ability to integrate and then say whole words after hearing parts (syllables and/or phonemes) of the word. An audiotape is used to present word parts in their proper order for each item.
	Incomplete words	.55	An audiotape subtest that measures auditory closure. After hearing a recorded word that has one or more phonemes missing, the subject names the complete word.
K-ABC	Not measured		—
SB IV	Not measured		—
WECH	Not measured		—
LONG-TERM RETRIEVAL (<i>Glr</i>)			THE ABILITY TO STORE INFORMATION AND FLUENTLY RETRIEVE IT LATER THROUGH ASSOCIATION.
WJ-R	Delayed recall: memory for names	.83	Measures the ability to recall (after 1–8 days) the space creatures presented in Memory for Names.
	Visual-auditory learning	.70	Measures the ability to associate new visual symbols (rebuses) with familiar words in oral language and to translate a series of symbols presented as a reading passage (a visual-auditory association task).
	Memory for names	.68	Measures the ability to learn associations between unfamiliar auditory and visual stimuli (an auditory-visual association task). The task requires learning names of a series of space creatures.
K-ABC	Not measured		—
SB IV	Not measured		—
WECH	Not measured		—
FLUID REASONING (<i>Gf</i>)			THE ABILITY TO REASON, FORM CONCEPTS, AND SOLVE PROBLEMS THAT OFTEN INCLUDE UNFAMILIAR INFORMATION OR PROCEDURES; MANIFESTED IN THE REORGANIZATION, TRANSFORMATION, AND EXTRAPOLATION OF INFORMATION.
WJ-R	Concept formation	.68	Measures the ability to identify and state the rule for a concept about a set of colored geometric figures when shown instances and noninstances of the concept.
	Analysis-synthesis	.59	Measures the ability to analyze the components of an incomplete logic puzzle and to determine and name missing components.
SB IV	Matrices	.61	Measures the ability to identify the response needed to complete a visual pattern matrix.

<i>Factor/ Battery</i>	<i>Subtest</i>	<i>Relative Factor Loading</i>	<i>Description</i>
K-ABC	Not measured		—
WECH	Not measured		—
PROCESSING SPEED (<i>G_s</i>)			THE SPEED OF AUTOMATIC COGNITIVE PROCESSING. IT IS OFTEN MEASURED BY CLERICAL-SPEED-TYPE TASKS, PARTICULARLY WHILE UNDER PRESSURE TO MAINTAIN FOCUSED ATTENTION.
WJ-R	Visual matching	.84	Measures the ability to locate and circle quickly the two identical numbers in a row of six numbers. The task proceeds in difficulty from single-digit numbers to triple-digit numbers and has a 3-minute time limit.
	Cross out	.62	Measures the ability to scan and compare visual information quickly. The subject must mark the 5 drawings in a row of 20 drawings that are identical to the first drawing in the row. The subject is given a 3-minute time limit to complete as many rows as possible.
WECH	Coding (digit symbol)	.58	Measures the ability to write quickly the digits associated with several simple drawings. The subject uses a reference key to the digit-drawing associations. The test has a 2-minute time limit in the WISC-R and a 90-second time limit in the WAIS-R.
K-ABC	Not measured		—
SB IV	Not measured		—

TABLE 2
Congruence of factor scores provided by four batteries of intelligence tests

<i>G_f-G_c</i> <i>Factor</i>	<i>Intelligence Battery</i>			
	<i>K-ABC</i>	<i>SB IV</i>	<i>WECH</i>	<i>WJ-R</i>
<i>G_{sm}</i>	Sequential processing	Short-term memory	—	Short-term memory
<i>G_c</i>	—	Verbal reasoning	Verbal comprehension	Comprehension-knowledge
<i>G_q</i>	—	Quantitative reasoning	—	Quantitative ability
<i>G_v</i>	Simultaneous processing	—	Perceptual organization	Visual processing
<i>G_a</i>	—	—	—	Auditory processing
<i>G_{lr}</i>	—	—	—	Long-term retrieval
<i>G_f</i>	—	—	—	Fluid reasoning
<i>G_s</i>	—	—	—	Processing speed

CONCLUSION

When factor-analytic studies are conducted with an adequate breadth and depth of measures, the studies provide useful information about the factorial structure of an intelligence battery. They also provide important information about the ability, or abilities, measured by the various subtests. *Gf-Gc* theory has provided a framework for interpreting the results of such studies. The factor analysis results reported in the literature for most published intelligence-test batteries have been based on studies including too few measures, usually only the subtests from a battery itself. As a result, the conclusions incorrectly describe the factor structure of the battery and the measurement focus of the separate subtests. Tables 1 and 2 present

results from a comprehensive factor-analytic study showing how example subtests and factor scores from contemporary intelligence-test batteries relate to the broad abilities of *Gf-Gc* theory.

BIBLIOGRAPHY

- WOODCOCK, R. W. (1990). Theoretical foundations of the WJ-R measures of cognitive ability. *Journal of Psychoeducational Assessment*, 8, 231-258. Provides a complete report of the factor studies referenced in this article.
- WOODCOCK, R. W. (1993). Example measures of *Gf-Gc* factors from four intelligence batteries. (Available from M/L/C., P.O. Box 161, Tolovana Park, OR 97145.) An expanded version of Table 1.

RICHARD W. WOODCOCK

G

GALTON, FRANCIS (1822–1911) If any figure in the history of the behavioral sciences can be characterized as an original genius, it is Sir Francis Galton. He never held an academic appointment, yet he is rightfully claimed to be the founding father of differential psychology, of psychometrics, and of behavioral genetics. His creative efforts presaged almost all of the major theoretical issues under investigation in these fields in the twentieth century. Few scientists have had such wide-ranging and lasting impact.

Galton was born on February 16, 1822, at a country estate near Birmingham, England, the last of seven children (three boys and four girls) in a Quaker family of wealth, culture, and privilege. His father was a prominent banker and civic leader. His mother was the daughter of Erasmus Darwin, famous in his day as a biologist, philosopher, and poet—he was also the grandfather of Galton's illustrious half-cousin, Charles Darwin. Galton's paternal grandfather, a largely self-taught scientist, was elected a Fellow of the Royal Society for his research in optics and astronomy.

Before Galton was 3 years old, it was evident that he was a prodigy, since he had learned to read. At age 4 he was reading simple books and writing letters to relatives in a style that would do credit to a child twice his age, a fact that led the psychometrician Lewis TERMAN to estimate Galton's childhood IQ at near 200. As a preschooler Galton was already reading Latin and Greek classics, but when he was sent off to a private

boarding school at age 8, he found it entirely uncongenial. He begged his parents to remove him. In his autobiography, written in his late 80s, Galton (1908) was still extremely critical of his early schooling: "I learnt nothing and chafed at my limitations. I had craved for what was denied, namely, an abundance of good English reading, well-taught mathematics, and solid science."

His parents wanted him to become a physician, following the footsteps of his eminent grandfather Erasmus Darwin. At age 16 Galton began the study of medicine, the first year at the Birmingham General Hospital, the second at Kings College Hospital in London, but only the basic-science aspects of medical education—chemistry, physics, mathematics, and physiology—captured his enthusiasm. He could envisage for himself a career in scientific research, not in treating patients. So at age 18, to prepare for a scientific career, he entered Cambridge University and majored in mathematics. A few months after Galton graduated from Cambridge, at the age of 21, his father died. Galton fell heir to an ample independent income from the family fortune and was freed from the need to earn a living for the rest of his life. He was immediately able to fulfill his long-held fascination with travel to remote places.

From this point on, Galton's long and immensely productive life can be divided into three main periods of intense activity devoted successively to exploration,

research on individual differences and genetics, and the promotion of eugenics.

TRAVEL AND EXPLORATION (1844–1864)

For twenty years, Galton traveled extensively in Europe and the Middle East and published books and journals about his experiences. It was in Africa, however, that he won distinction as an explorer, geographer, and anthropologist. He extensively explored and mapped little-known regions of tropical South Africa and studied the physical and social characteristics of the people. One of his books based on his African explorations, *The Art of Travel* (1855), was a long-time popular bestseller in its day and went through nine editions. Great Britain's Royal Geographical Society awarded Galton its gold medal for his explorations, and his name is included among the famous explorers in British history—engraved on the granite facade of the Royal Geographical Society's headquarters in London.

During this period Galton also became engrossed by meteorology, to which he made original contributions. He was a pioneer in weather mapping and was the first to write weather reports for a daily newspaper, *The Times* of London. Also, he formulated a theory of cyclones and discovered the anticyclone. In developing better methods for predicting the weather by taking simultaneously into account a number of predictive indices, he invented a graphical form of what later, in algebraic form, became known as multiple-regression analysis.

INDIVIDUAL DIFFERENCES, STATISTICS, AND INHERITANCE (1865–1899)

This was the most productive period of Galton's career in science. His compulsive curiosity, mechanical ingenuity, mathematical bent, and theoretical inventiveness resulted in original contributions to meteorology, composite photography, graphology, fingerprint classification, anthropometry, sociology, education, genetics, psychometrics, and statistics. Whole chapters in books on the history of some of these fields have been devoted to his contributions (e.g., Stigler, 1986). Galton's accomplishments were recognized formally by

his being awarded virtually every honor and distinction available to the most eminent personages of that period—knighthood, honorary degrees, fellowship in the Royal Society, and awards from many scientific societies.

In Galton's time, psychology was dominated by the philosophy and methodology of Wilhelm Wundt's psychological laboratory in Leipzig, Germany, which was mainly concerned with discovering the general laws of sensation, perception, and other mental processes, as measured by a variety of laboratory instruments (from which originated the phrase "brass instrument psychology"). Individual differences in these processes were of no interest to Wundt and his followers, however, who regarded this source of variation merely as a "nuisance" variable, to be minimized by averaging large numbers of measurements. Although Galton knew of the studies being done in Wundt's laboratory, he showed little interest in Wundt's theoretical aims, but he was impressed by Wundt's techniques for objectively measuring sensory-motor functions. Meanwhile, Galton's encounter with Darwin's *The Origin of Species* (1859), which he claimed had a greater impact on his thinking than any other book he had ever read, inspired his passion for understanding human evolution and human variation in both physical and mental traits. For the rest of his life, Galton was engrossed in studies of the biological basis of individual differences, proposing empirically testable theories, collecting enormous amounts of data, and inventing methods of measurement and statistical analysis that have mostly endured to the present. The individual differences that Wundt regarded as a nuisance variable were viewed by Galton as a paramount phenomenon for scientific study.

Galton well knew, of course, that natural variation among individuals of the same species is a crucial pillar of Darwin's theory of evolution. Individual differences in particular characteristics, based on genetic influence, constitute the raw material on which the process of natural selection works to effect changes in the course of evolution. A science of individual differences in human characteristics, Galton believed, would have immense social significance bearing on the future of human welfare. His interest focused mainly on individual differences in those mental abilities and personality traits he thought were most related to socially valued

achievements. Thus, he launched the fields of differential psychology and its methodological basis, psychometrics (which measured those traits). Together they constitute one of “the two disciplines of scientific psychology” (the other is experimental psychology), as described by Lee Cronbach (1957) in his frequently cited presidential address to the American Psychological Association.

Galton’s contributions in differential psychology are in some 65 of his more than 300 scientific publications. The fruits of his research on the topics most germane to the theory and measurement of intelligence were incorporated in two of his best-known books, *Hereditary Genius* (1869) and *Inquiries into Human Faculty and Its Development* (1883). His specific contributions in this domain are perhaps best summarized under the following headings of the contemporary fields they have most influenced.

Psychometrics and Statistics. Galton was never interested in the methodology of measurement and statistics for its own sake. All of his methodological contributions were merely incidental to the substantive questions that interested him. He quantified virtually everything. Galton’s favorite motto was, “Whenever you can, count.” He promoted the idea of objective measurement and quantitative analysis of data, whether by counting, ranking, or by true measurement. He held that objective measurement and the mathematical treatment of quantitative data were essential for a science of human variation; he applied this philosophy to most physical and mental characteristics that were within his power to count, rank, or measure. Thus, he originated a number of statistical and psychometric concepts familiar to present-day researchers.

One subject that interested Galton, for its obvious genetic implications, is the degree of resemblance between parents and offspring in various physical and mental traits; his efforts to research this question led to his methodological contributions best known today. As recently as the late nineteenth century no means existed for expressing, in an exact quantitative way, the degree of resemblance or association between two variables. No means existed by which one could precisely answer whether offspring are more similar to their parents in Trait X than in Trait Y. Galton began his study of parent-offspring resemblance in a precisely

measurable trait—stature. He measured nearly 1,000 young adults and their fathers and mothers. (The height measurements were adjusted to remove sex differences by multiplying all female heights by 1.08.) The total range and shape of the distribution of height, as well as the overall mean height, were nearly the same for parents and offspring. It is obvious that individual offspring are seldom the same height as their parents, so Galton pursued how to represent precisely in quantitative terms the degree of resemblance between parents and offspring. He solved this by making a bivariate plot of the parent-offspring measurements, forming what we now call a *scatter diagram*.

He discovered that for his height project the data points formed an upward-sloping ellipse, which prompted the next step. After plotting the median height of all offspring who fell within each one-inch interval of parental height, he drew a single straight line closest to the center of this array of medians. The line was expressed mathematically as a simple linear equation, and its slope (or “regression coefficient”) would serve as a precise measure of the degree of parent-offspring resemblance. Because the distribution of height was almost the same in parents as in offspring, when the axes of the scatter diagram were reversed and parents’ medians were plotted within each 1-inch interval of offsprings’ height, virtually the same line fitted the array of medians. The *slope* of the best-fitting line, which was the same in both cases, Galton referred to as the coefficient of co-relation (later spelled correlation). Thus was invented the concepts we now know as linear *regression* (i.e., the best-fitting straight line to the array of means in a bivariate scatter-plot) and *correlation* (i.e., the slope of the regression line when both variates are measured on the same scale and have the same variance.) Galton applied his methods to “parent” and “offspring” generations of size in sweet peas, with a result very similar to that for human height.

Galton was also interested in the degree of association between different traits measured on different scales and had to invent a way to express the correlations between them. This led to his invention of standardized measurements, that is, rescaling the original measurements to a common scale for all variables, known in psychometrics as *standardized scores*. The modern formulation of correlation, however, is attrib-

utable to Karl Pearson (1857–1936), a mathematician who greatly admired Galton and gave mathematically rigorous formulations to many of Galton’s intuitive statistical conceptions. Pearson, known as the “father of mathematical statistics,” was also Galton’s chief disciple and a pioneer of *biometrical genetics*. Pearson wrote the most comprehensive biography of his hero: *The Life, Letters and Labours of Francis Galton* (3 vols., 1914–1930).

Other methods invented and introduced by Galton include: the scaling of mental test scores in terms of their percentile ranks within a specified population; the use of rating scales for nonmetric traits; scaling test scores and ratings in terms of the normal curve; use of the median and geometric mean as measures of central tendency in markedly skewed distributions; scaling skewed data by the lognormal distribution; and the ogive. Also, some of his ingenious quantitative analyses of data—though largely intuitive and lacking rigorous mathematical derivation—are the conceptual forerunners of *multiple regression*, *multiple correlation*, the *analysis of variance*, and *factor analysis*. The systematic development of these ideas were accomplished by mathematicians, mainly Karl Pearson and Ronald Fisher.

Behavioral Genetics. Galton was born the same year as Gregor Mendel (1822–1884); both of them, unknown to each other at the time, were pioneers in genetics, experimenting with peas as a means for investigating “natural inheritance.” Both independently arrived at the conclusion that heredity is particulate, that is, traits are inherited via discrete elements (later called *genes*) that pass (for the most part) unchanged from generation to generation. Mendel made the greater contribution, since he discovered the fundamental laws of heredity—segregation, independent assortment, and dominance/recessiveness—whereas Galton never did. Mendel’s success can be attributed to his propitious choice of phenotypes for his breeding experiments—several visible characteristics of peas (color [green or yellow] and form [smooth or wrinkled]) that were discrete, hence countable, and were determined (fortuitously) by single genes with no genetic linkage to other characteristics. Galton did his breeding experiments with peas merely as a convenient way of discovering things about genetics that might apply to human heredity, which are continu-

ously distributed in the population, such as height and mental ability. He focused his study on a continuous trait in peas—their size (or weight)—which we now know is determined polygenically and to some extent environmentally. His data were too complicated therefore to reveal the fundamental rules discovered by Mendel. Nevertheless, Galton’s studies revealed some important facts about polygenic inheritance. He discovered that the distribution of a polygenic trait conforms approximately to the normal curve in each generation, even when the parental generation is selected so as to have a markedly nonnormal distribution, and that, in the absence of selection, the variance of the trait was constant across generations.

Galton believed his most important discovery to be the phenomenon he first termed “reversion to the mean,” and later, “regression to the mean.” On finding the same phenomenon with respect to both physical stature and mental ability, he dubbed it *The law of filial regression to mediocrity*. This “law” simply quantified the observation that, in a given trait, offspring, *on average*, do not deviate as much from the mean of the population as do their parents. Of course, such regression toward the mean is simply a corollary of the imperfect correlation between parents and offspring—which is the more basic phenomenon. Hence, parent-offspring regression works in both directions: parents, *on average*, deviate less from the population mean than do their children. Although Galton’s “law of regression” can be explained in terms of genetic theory, environmental factors and measurement error may also contribute.

Galton’s “law of ancestral inheritance” has not survived in modern genetics. Neither Galton nor anyone else of that time fully understood the genetic mechanisms underlying “regression to the mean,” which actually involves nonadditive effects (now known as *dominance* and *epistasis*) due to interactions among genes. In each generation the genes are simply shuffled, so to speak, and their interaction effects are redistributed at random. Thus, a parent who is exceptional in some trait because of propitious gene interactions may have an unexceptional child; and, for the same reason, an exceptional individual may have quite unexceptional parents.

Consider next Galton’s most famous work, *Hereditary Genius* (1869), about which Charles Darwin re-

marked, “I do not think I ever in all my life read anything more interesting and original.” It was the first attempt to study scientifically the inheritance of mental ability. As there were no validated intelligence tests at that time, Galton used as his criterion of mental ability the attainment of eminence in intellectual achievements. He began with a sample of some 400 historical figures—scientists, writers and poets, composers, statesmen, divines, judges, and the like—for whom extensive biographical data could be found in libraries (e.g., Aristotle, Newton, Goethe, Beethoven, Napoleon Bonaparte, Richelieu, and Disraeli). He labeled these individuals “illustrious.” From biographical sources, he traced their direct-line ancestors and descendants as well as their collateral relatives (brothers, uncles, nephews), and determined the percentages of these groups who were at least distinguished enough by their achievements to be found in biographical directories. He labeled this level “eminent.” Galton’s two main findings were (1) a much higher probability of eminence among the genetic kinships of the illustrious than is found for persons selected at random from the general population; (2) the percentage of eminent persons decreases in a regular stepwise fashion the farther the degree of kinship is removed from the illustrious. Galton performed the same kind of analysis on champion athletes and obtained a highly similar result for athletic distinction.

Galton also introduced the *adoption method* for studying the relative effects of heredity and environment—“nature and nurture”—to use the phrase for which he is generally credited but which comes from Shakespeare (*Tempest*, IV, i). Noting that it had been customary in the past for popes to rear adopted sons, Galton found that their adopted sons, despite environmental advantages comparable to the natural sons of other illustrious men, did not show as adults anywhere near the same level of distinction as the biological descendants of the illustrious and eminent.

From his finding that the major distinctions of eminent relatives were often in a variety of fields (mathematics, literature, musical composition), Galton also concluded that distinguished intellectual achievements of any kind, or at least their hereditary component, are due to a *general* mental ability, which can be channeled by circumstance or interest into almost any

kind of intellectual endeavor. This notion closely resembles the modern concept of *fluid ability* (and its associated “investment” theory of intelligence) formulated by R. B. Cattell in 1943.

The *twin study* method, which has figured most prominently in behavioral genetics, is also attributable to Galton. He was the first to note the importance of twins for determining the relative effects of heredity and environment, and particularly the significance of there being two distinct kinds of twins, *identical* and *fraternal*, or monozygotic (MZ) and dizygotic (DZ). Galton collected information on ninety-four sets of twins and studied their resemblances in many physical and behavioral characteristics, even their history of illnesses. He noted in examining the frequency distribution of *differences* between twins in various traits that the distributions are typically *bimodal*, indicating that there are two distinct types of twins. Since MZ twins are genetically identical and DZ twins are genetically no more similar than ordinary full siblings, Galton realized that a comparison of the correlation (r) between MZ twins with the correlation between DZ twins on a particular trait would indicate the degree to which genetic factors influence individual differences in the trait. Galton’s insight is the basis for a commonly used formula in modern genetics for estimating the *heritability* (h^2) of a trait, in other words, the proportion of phenotypic variance in the trait attributable to genotypic variance: $h^2 = 2(r_{MZ} - r_{DZ})$. Summarizing his study of twins, Galton wrote:

There is no escape from the conclusion that nature prevails enormously over nurture when the differences in nurture do not exceed what is commonly to be found among persons of the same rank of society and in the same country. My fear is, that my evidence may seem to prove too much, and be discredited on that account, as it appears contrary to all experience that nurture should go for so little (1883/1907, p. 172).

Theory and Measurement of Mental Ability. Galton never presented a formalized theory of intelligence, but it is clear from several of his publications, particularly *Inquiries into Human Faculty and Its Development* (1883/1907), that his conception of it can be summarized as *innate, general, cognitive ability*. The specification “cognitive” distinguishes it from the two

other main aspects of mind recognized by Galton, the *affective* and the *conative*. He theorized that this general ability is a Darwinian fitness characteristic—hereditary—which developed through natural selection in the course of human evolution. He held that the distribution of intelligence in the population, like that of many hereditary physical traits, conforms to the normal, bell-shaped curve. His conception of ability as a normally distributed continuous trait represented a break with the typological thinking of his contemporaries; they viewed both genius and mental deficiency as distinct types of intelligence, separate from the general run, not the upper and lower extremes of a continuous distribution. Besides his hypothesis of individual differences in a *general* ability that is involved in all intellectual activity, he also recognized the existence of *special* abilities and talents but attributed less importance to them than to general ability in accounting for individuals' lifetime achievements.

Because all the contents of the mind enter through the sensory system and all expressions of mind depend on the effector system of motor nerves and muscles, Galton, arguing from evolutionary concepts, hypothesized that precise laboratory measurements of (1) human sensory acuity and discrimination and of (2) reaction times to the onset of a visual or auditory stimulus would afford an objective assessment of individual differences in the biological basis of general mental ability.

Galton invented an extensive battery of devices—various tests of sensory discrimination, reaction time, and memory span—with which to measure individual differences (described in his *Inquiries into Human Faculty*). In his Anthropometric Laboratory in the South Kensington Natural Science Museum, he and two assistants administered this battery of sensory-motor tests, along with a number of physical measurements, to nearly 10,000 people. He found that the sensory and reaction-time measurements seemed to show only very small differences between those classified into ability levels based on education and occupation; this was disappointing and Galton did not pursue this avenue further. Others did but with little more success at the time. It turns out, however, that Galton's original hypothesis was essentially correct; this fact remained unrealized in his time because of the generally

low reliability of much of his data—especially that on reaction time (with a reliability coefficient of only .18)—and because appropriate statistical techniques had not yet been invented for determining the significance of his findings. When modern statistical techniques, such as the analysis of variance and multiple regression, were applied to Galton's original massive data, they confirmed his observations—that great overlap exists for the score distributions of the various occupational and educational categories and that the *average* differences (in the predicted direction) between these various categories appear almost negligibly small. It was also found that most of the average differences are statistically highly significant ($p < .001$), which in Galton's time could not be ascertained (Johnson et al., 1985). Today we know that the newer measurement techniques yield high precision and reliability for some of the simple variables of interest to Galton; they reveal low to moderate correlations with scores on modern psychometric tests of intelligence. Many of Galton's intuitive hypotheses have proved amazingly fruitful, although in his time the technical means for properly testing them was lacking.

EUGENICS (1900–1911)

Galton dedicated the last decade of his life to the advancement of *eugenics*, a term he coined for the study of genetics for the general betterment of the human species. In his autobiography (1908), written at age 88, he summarized his vision of eugenics:

Man is gifted with pity and other kindly feelings; he has also the power of preventing many kinds of suffering. I conceive it to fall well within his province to replace Natural Selection by other processes that are more merciful and not less effective. This is precisely the aim of Eugenics. . . . I take Eugenics very seriously, feeling that its principles ought to become one of the dominant motives in a civilized nation, much as if they were one of its religious tenets (pp. 322–323).

He even wrote a utopian novel (*Kantsaywhere*) based on eugenics. A man of action, Galton founded a number of institutions to advance genetics and eugenics and used most of his personal fortune for perpetual

endowments. (Although married for 44 years, he had no children or heirs.) He endowed the Galton Laboratory of Genetics at the University of London, the first of its kind; it remains a leading center of research in human genetics. He endowed a Professorship in Eugenics (now Genetics) at the University of London, which has been occupied by such luminaries as Karl Pearson, Sir Ronald A. Fisher, and Lionel S. Penrose. Galton founded and endowed two prestigious journals, *Biometrika* and *The Annals of Human Genetics*, which are still published. He also founded the Eugenics Society (recently renamed The Galton Institute), which publishes its own journal (*Journal of Biosocial Science*) and an annual symposium series on a wide variety of topics related to human genetics and eugenics (now more commonly referred to as social biology). He died one month short of age 90, on January 17, 1911. Few scientists have left such a diverse and lasting influence.

(See also: HERITABILITY; TWIN STUDIES OF INTELLIGENCE.)

BIBLIOGRAPHY

- BORING, E. G. (1950). *A history of experimental psychology*. (2nd ed.). New York: Appleton-Century-Crofts. Chapter 20.
- BURT, C. (1962). Francis Galton and his contributions to psychology. *British Journal of Statistical Psychology*, 15 (Part 1), 1–49.
- CRONBACH, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist*, 12, 671–684.
- FORREST, D. W. (1974). *Francis Galton: The life and work of a Victorian genius*. New York: Taplinger.
- GALTON, F. (1869). *Hereditary genius*. London: Macmillan.
- GALTON, F. (1883; 2nd ed., 1907). *Inquiries into human faculty and its development*. New York: Dutton. (Reprinted 1973 by AMS Press, New York.)
- GALTON, F. (1889). *Natural inheritance*. London: Macmillan. (Reprinted in 1973 by AMS Press, New York.)
- GALTON, F. (1908). *Memories of my life*. London: Methuen.
- JENSEN, A. R. (1987). Individual differences in mental ability. In J. A. Glover & R. R. Ronning (Eds.), *Historical foundations of educational psychology*. New York: Plenum. Pp. 61–88.
- JOHNSON, R. C., MCCLEARN, G. E., YUEN, S., NAGOSHI, C. T., AHERN, F. M., & COLE, R. E. (1985). Galton's data a century later. *American Psychologist*, 40, 875–892.

KEYNES, M. (Ed.). (1993). *Sir Francis Galton, FRS: The legacy of his ideas*. London: Macmillan.

STIGLER, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Cambridge: Harvard University Press.

ARTHUR R. JENSEN

GENDER DIFFERENCES IN INTELLECTUAL ABILITIES

The questions of whether, when, and how much females and males differ in their intellectual abilities have resulted in an enormous amount of hard feelings and controversy among researchers who have collected mountains of data in an attempt to answer these questions. Some tentative answers have emerged from all the data along with a clearer definition of the questions that still need to be answered. Although there is still a great deal to be learned about gender differences in the ability to think, learn, and remember, a gender-differentiated pattern of intellectual abilities has been found with females, on the average, performing better on a wide variety of cognitive tasks that involve the rapid retrieval of information from memory. Males, on the average, perform better on cognitive tasks that involve maintaining and manipulating information in short-term memory. It is also clear that a satisfactory explanation of these differences will involve an interaction of psychological, biological, and social variables. In interpreting these conclusions, it is important to keep in mind the fact that reports of “average” differences can be misleading because there is considerable overlap between the sexes in all intellectual areas.

WHAT ARE INTELLECTUAL ABILITIES?

Most psychologists think that intelligence is made up of numerous component abilities that are measured with performance on intellectual tasks. Some of these tasks are closely related, such as the ability to spell well and the ability to detect spelling errors in printed text; other intellectual tasks are more independent, such as the ability to hold an image in memory while deciding what it would look like if it were rotated in space (mental rotation) and the ability to generate synonyms for words. There have been numerous attempts

to categorize tasks that rely on similar processes in order to identify intellectual abilities and the tasks that can be used to measure them.

Intellectual abilities have been identified empirically with a mathematical procedure known as factor analysis. In a typical factor-analysis study, subjects participate in a variety of tasks. The results of these tasks are then subjected to a factor analysis that shows which tasks are similar via a process known as factor loadings. A more contemporary variation of factor analysis is called confirmatory factor analysis. It is a more conservative and theory-driven procedure in which the researchers specify in advance of the analysis those tasks that they believe are measuring the same ability. They are then able to see how well the data fit the theoretical model they have specified.

Factor analyses have shown that there are at least three different intellectual abilities that psychologists frequently study—verbal, quantitative, and spatial abilities. The emergence of these three separate factors suggests that they are relatively independent abilities, which means that an individual's score on one of these factors is unrelated to his or her score on the other factors. Examples of tasks that are used to assess verbal abilities include tests of grammar, spelling, reading comprehension, the rapid generation of synonyms, and verbal analogies. Quantitative abilities are indexed by tests of computational arithmetic, algebra problems, probability, mathematical word problems, and more advanced mathematics such as topology and calculus. Spatial abilities are measured within mental rotation problems, figures embedded with the borders of other figures, map reading, moving spatial arrays, and mazes.

Although the division of intellectual abilities into three broad categories has been a useful way of conceptualizing these abilities, it is now clear that this distinction is not the best way of organizing the results of gender-differences research. Each of these abilities is multivariate, and gender differences are found on some, but not all, tasks within each of these broad ability headings. A more fruitful way of understanding gender differences is to examine the underlying cognitive process involved in performing the tasks rather than focusing on whether the task involves the use of language, the use of numbers, or the use of spatial displays.

TESTS THAT SHOW GENDER DIFFERENCES

When differences are found on tests of verbal abilities, the results usually favor females, with males scoring higher than females only on tests of verbal analogies. The largest differences favoring females are found on tests of word fluency (the ability to generate words both in isolation and in a given context) and on vocabulary tests in which subjects provide synonyms for target words. Effects are also large at the low end of the verbal-abilities distribution with males comprising the overwhelming majority of those with stuttering problems, some forms of retardation, and dyslexia. Surprisingly, males outscore females slightly, by approximately 13 points on the verbal portion of the Scholastic Aptitude Test (SAT-V), the most frequently used test for making decisions about college admissions. Although this result may seem contrary to the general conclusion that females tend to excel in most verbal tasks, the SAT-V is heavily weighted with analogies, and analogies are the only verbal task that favors males.

Visual-spatial abilities are comprised of at least four component processes—spatial perception, which requires subjects to locate the horizontal or vertical in a display while ignoring distracting information; mental rotation, which includes the ability to image how an object would look if it were rotated in space; spatial visualization, which is a catchall category for tasks such as finding figures that are embedded in the contours of larger figures; and spatiotemporal tasks, which involve making judgments about moving figures such as determining the time of arrival for a projectile. Of these four types of visual-spatial tasks, males tend to perform more quickly and more accurately on those that involve manipulations in short-term memory. The largest differences occur with mental rotation, spatial perception, and spatiotemporal tasks. Gender differences on spatial-visualization tasks are so small that they are essentially nonexistent.

A mixed pattern of results is also found with quantitative measures. Females score higher on computational arithmetic tests during the early elementary school years, with no differences in middle school; males score higher on quantitative tests in later high

school years. Males, as a group, also tend to score higher on tests of mathematical problem solving, which are usually administered in the later high school years. The largest differences favoring males are found on the mathematics portion of the Scholastic Aptitude Test (SAT-M), with an approximate 50-point lead for male college-bound high school seniors over their female counterparts.

AN UNDERLYING COGNITIVE PROCESSES APPROACH

This somewhat mixed pattern of gender-differentiated results can be more meaningfully explained by considering what subjects are doing when they take these tests. Tests that consistently show a female advantage include computational arithmetic, the production of fluent speech, and the ability to provide synonyms and solve anagrams. All these tasks involve rapid access to and retrieval of information from memory. By contrast, the tasks at which males excel include solving verbal analogies, making judgments about moving figures, and solving mathematical problems. All these tasks involve maintaining and manipulating mental representations. Thus, another way of categorizing gender differences in intellectual abilities is to group them according to the cognitive operations that are used in performing different intellectual tasks.

ETIOLOGY OF GENDER DIFFERENCES IN INTELLECTUAL ABILITIES

The "Why" Question. Listing ways in which females and males differ on the average is easier than determining why these differences are sometimes found. It is clear that there are no simple answers for questions about complex issues and that psychological, biological, and social factors are involved in creating and maintaining these differences.

Biological Variables: The Sexually Dimorphic Brain. Males and females differ in many ways in addition to the obvious differences in the shape and function of their reproductive organs. There are also gender differences in the structure and organization of some portions of the brains. An organ is labeled sexually dimorphic if it occurs in two forms that differ as

a function of sex. The brain is one such organ. Some of the differences involve portions of the brain that are involved in sexuality and in menstruation (the hypothalamus); others are more likely involved in intellectual processes. Recent research has found that one section of the broad band of neural fibers that connect the two hemispheres of the brain (the corpus callosum) differs in size for males and females. This is a potentially important finding because the two hemispheres are somewhat dominant or lateralized for different intellectual functions, and communication between the two halves of the brain is believed to be mediated by these fibers. Lateralization means that one hemisphere is more or less specialized or proficient in its ability to handle specific tasks such as verbal and visual-spatial tasks.

In addition to differences in the size of these neural fibers, some researchers have found evidence that males and females have different patterns of lateralization. A popular theory of gender differences in lateralization is known as the cognitive-crowding hypothesis. According to this hypothesis, most males have verbal abilities lateralized in the left hemisphere and visual-spatial abilities lateralized in the right hemisphere. Females also have visual-spatial abilities lateralized in their right hemisphere, but neural space devoted to verbal abilities is found in both hemispheres. If this theory is correct, it is possible that females excel in verbal abilities because they have more neural area devoted to these abilities, and visual-spatial abilities suffer for females because visual-spatial abilities get crowded out by having to share space with verbal abilities. Although there is a diverse body of evidence that supports this theory, many researchers believe that the support is weak. This is a highly controversial theory among brain researchers. There is also empirical support for a related theory of gender differences in lateralization. Some investigators believe that brain differences between the sexes are more frequently found within each hemisphere, with different intrahemisphere areas of the brain specialized for intellectual tasks in females and males.

The brains of females and males also differ in portions of the thickness of the outer brain covering (the cerebral cortex) and in the structure of neurons in some areas of the brain, although it is difficult to un-

derstand how or if these differences affect the thinking process.

Sex Hormones. Another important way in which the sexes differ is in the relative concentrations of hormones that are commonly called sex hormones. Androgens, most typically testosterone, are usually thought of as male hormones, and estrogen and progesterone are usually thought of as female hormones, even though all these hormones are found in detectable amounts in all normal females and males. These hormones are secreted by the gonads (testes in males and ovaries in females), adrenal glands, and other structures such as fat. Like all hormones, they travel through the blood stream so that they can affect target organs. Sex hormones are critically important during prenatal development because the genitals develop in either a male or female direction under the control of these hormones. The male hormones are responsible for the development of male genitals and reproductive organs and the absence of male hormones is implicated in the development of female genitals and reproductive organs. Prenatal hormones are also involved in brain development, with both female and male hormones influencing the way the brain is formed. Thus, sex hormones play a role in directing brain morphology, but the specific nature of this role is still unknown.

Sex hormones are also critically important at puberty as they underlie most of the developmental differences between boys and girls. Research has shown that these hormones also play a role in the development of intellectual abilities at puberty. Research with males who had a testosterone deficiency found that a minimal amount of testosterone is needed at puberty in order for them to exhibit visual-spatial abilities. Additional testosterone administered later in life could not reverse the severe deficit in visual-spatial abilities that was caused by the testosterone deficiency at puberty. The role that testosterone plays in normal males and females at puberty is unknown as is the role of other sex hormones at puberty.

Psychosocial Variables. There is no doubt that environmental and social variables differ systematically for females and males in our society. Sex-differentiated practices and messages are so prevalent and so ingrained in western society that they are frequently underestimated. Gender-differentiated practices begin at birth with pink and blue booties issued as a way of

telling diapered children apart in the newborn nursery. Toy stores are often arranged into girl and boy sections with toys that train visual-spatial skills (building blocks, dominoes, Legos, tinker toys) in the boys' section and playhouse toys, dolls, and books in the girls' section. It is easy to see how these experiences could affect the development of intellectual abilities in a gender-differentiated manner.

Sex-role stereotypes are those beliefs about behaviors and dispositions that are appropriate for members of each sex. These sex-role stereotypes influence life experiences, which in turn can affect the development of intellectual abilities. For example, mathematics is a male-stereotyped intellectual domain. Females receive less encouragement to excel in mathematics and females report that mathematics is not as important for their career goals as it is for males. Given these attitudes, it is not surprising that females do not score as high as males on advanced tests of quantitative abilities. Sex-role stereotypes affect other intellectual abilities as well. Girls receive more books than boys and engage in more doll play, a type of play that has been found to be highly verbal. Similarly, visual-spatial abilities can be developed through play with spatial toys and with subtle and overt messages about the sorts of activities that are acceptable for males and females.

A Psychobiosocial Model of Intellectual Development. It is clear that both biological and psychosocial variables have important influences on the development of gender-differentiated patterns of intellectual abilities. However, the effects of these two types of variables are not independent. A three-part model in which psychological, biological, and social variables exert mutual influences on each other is a more accurate depiction of the underlying processes.

The effect of biological variables on psychological and social variables is sometimes called the bent-twig hypothesis. The name of this hypothesis comes from an old adage that states that as the twig is bent, so the tree shall grow. According to this hypothesis, people adjust and select their life experiences so that they are in accord with their natural propensities and strengths. For example, if males are somewhat better at visual-spatial tasks for biological reasons, they are more likely to seek activities that are spatial in nature, and the gender difference will become even larger than it initially was. Similarly, if females are, on the average, bet-

ter at reading, then they will read more often and further develop their reading skills relative to males. In this way, biological factors affect psychosocial ones.

The reciprocal effect of psychosocial factors on biology has been recognized more recently by researchers. For example, brain structures and hormone concentrations respond to life experiences. Experimental work with nonhuman mammals has shown that an enriched environment will result in more complex neurons in the brain and in increased thickness in the cerebral cortex. Furthermore, the effects of social experiences on brain structures can be seen well into old age. Although we do not have direct experimental evidence of the same effect with humans (and cannot perform this research because of ethical considerations), the same sort of environmental influences on brain structures are likely. Similarly, stress and other life experiences are known to affect the secretions of hormones including the sex hormones. In this way, social variables can shape the biological structures that underlie cognition.

Psychological variables also include sex-role identification and adherence, self-confidence, and perceived abilities. If females are socialized to believe that they have less ability in mathematics, for example, it is likely that they will not take advanced mathematics course work or work as hard to learn advanced mathematics concepts. Thus, psychological variables will determine social interactions, which can affect brain structures, and so on. Psychological, biological, and social variables are inextricably entwined; it is not possible to assess the effects of any one of these influences without also implicating the other two. This seamless web of variables is needed to explain gender-differentiated patterns of intellectual abilities.

(See also: SEX CHROMOSOMAL ABNORMALITIES.)

BIBLIOGRAPHY

- BAKER, M. A. (Ed.). (1987). *Sex differences in human performance*. Chichester, England: Wiley.
- BENBOW, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. *Behavioral and Brain Sciences*, 11, 169–232.

EAGLY, A. H. (1987). *Sex differences in social behavior: A social-role interpretation*. Hillsdale, NJ: Erlbaum.

HALPERN, D. F. (1992). *Sex differences in cognitive abilities* (2nd ed.). Hillsdale, NJ: Erlbaum. Contains an extensive discussion and a complete list of references.

DIANE F. HALPERN

GENERAL APTITUDE TEST BATTERY

During the years 1934–1942, the U.S. Department of Labor developed aptitude tests to predict job performance for 100 specific occupations. During the years 1942–1945, the department invited a blue ribbon panel of the top experts in industrial psychology and measurement to help create a battery of employment tests to cover all of the 100 occupations and many more. The subsequent battery became the General Aptitude Test Battery (GATB). The GATB was then and is now a state-of-the-art broad spectrum test measuring all of the aptitudes most useful in predicting job performance. The construct validity of each test is high because the GATB either used or introduced the tests that are among the best measures of each aptitude. The reliability of each test is about as high as can be obtained in the time available for employment testing. The basic characteristics of the test and norms for a wide variety of groups can be found in the manual (United States Employment Service, 1970). The office within the Department of Labor that was responsible for the development and validation of the GATB was for many years called the United States Employment Service (USES); the name is now Job Service.

The GATB measures nine aptitudes that can be grouped in three measures of as many general abilities each. There are three measures of general cognitive ability: (1) intelligence; (2) verbal aptitude; and (3) numerical aptitude. There are also three measures of general psychomotor ability: (1) motor coordination; (2) finger dexterity; and (3) manual dexterity. There are also three measures of general perceptual ability (perhaps better called *general perceptual speed*): (1) spatial aptitude; (2) form perception; and (3) clerical perception. John Hunter (1983a) found that the prediction of job performance was determined largely by general abilities rather than by specific aptitudes, although there are a few exceptions connected with spatial ap-

titude and jobs using the graphic arts. Once cognitive and psychomotor abilities are considered, perceptual ability is largely redundant, though it does contribute to the prediction of performance for industrial-setup workers.

Hunter (1986) reviewed the theory behind the GATB along with massive supporting data. Hundreds of studies show that the GATB abilities measure how well a person can learn the knowledge and skills that determine job performance.

Since 1947, the Job Service has done over 700 validation studies testing the extent to which the GATB predicts job performance and employment training success. These data were pulled together using meta-analysis by Hunter (1983a, 1983b). His work was critiqued by a National Academy of Science panel (Hartigian & Wigdor, 1989), which found the GATB to be as good as any test used for predicting job performance. The panelists found that the GATB predicts job performance better than existing alternatives such as interviews and reference checks.

Hunter and colleagues (1984, 1987) reviewed the research done by the Job Service along with a great deal of other corroborating work on ability and performance. For the purpose of selecting employees who must be trained after hiring, the GATB predicts job performance far better than any known alternative (except other directly congruent ability tests). For the purpose of selecting employees who have already worked at the job, the GATB predicts performance only slightly less well than a job-specific content-valid job-knowledge or work-sample test.

The prediction of performance ratings varies with job complexity. As job complexity goes up, cognitive ability predicts performance better, and psychomotor ability predicts performance less well. The optimal combination of the two varies with complexity, but the overall prediction using both varies little—a multiple correlation varying from .49 to .58.

Cognitive ability predicts training success with high validity across all levels of job complexity, the average correlation being .62. However, psychomotor ability is much more valid for jobs of low complexity than for jobs of high complexity. It makes a moderate contribution to the prediction of training success in low-complexity jobs and none to jobs of medium or higher

complexity. There is evidence suggesting much higher validity predicting the learning rate of developmentally disadvantaged workers, as is consistent with extrapolations from the lowest level of complexity considered in Job Service studies.

Extensive research has been done to check the fairness of the GATB in measuring the ability of various minority workers. The GATB is a fair measure of ability for blacks as well as whites, for Hispanics as well as others, for women as well as men. Some of the tests must be used with care in assessing the ability of the handicapped.

Many persons believe that ability has little to do with job performance. They believe that job performance should be better predicted by measures of personality and motivation. These beliefs are not consistent with empirical findings (Schmidt & Hunter, 1992). For example, a close analysis of the findings on nineteen military specialties (Hough et al., 1990) shows that among all the main dimensions of personality, only measures of conscientiousness correlate with objectively measured job performance, and the correlation is only about .10. By contrast, the correlation between ability and objectively measured performance in their data is .68.

There is a higher correlation between personality and subjective performance ratings by supervisors. M. Barrick and M. Mount (1991) did a meta-analysis of hundreds of studies relating personality to subjective performance ratings. They found evidence of relevance only for conscientiousness and not for other personality traits. They found an average correlation between conscientiousness and subjective ratings that is considerably higher than the correlation for objective measures of performance, about .26 rather than .10. In the military studies by Hough and associates (1990), the correlation between conscientiousness and subjective performance ratings was about .42. The meta-analysis findings on conscientiousness (Ones, Viswesvaran, & Schmidt, 1993) are closer to that higher figure, an average correlation being .35. This situation is in comparison to an average correlation of .57 between the GATB aptitudes and subjective performance ratings.

(See also: APTITUDE TESTS; JOB PERFORMANCE.)

BIBLIOGRAPHY

- BARRICK, M. R., & MOUNT, M. K. (1991). The Big Five personality dimension and job performance: A meta-analysis. *Personnel Psychology*, 44, 1–26.
- HARTIGAN, J. A., & WIGDOR, A. K. (1989). *Fairness in employment testing*. Washington, DC: National Academy of Science.
- HOUGH, L. M., EATON, N. K., DUNNETTE, M. D., KAMP, J. D., & MCCLOY, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Personnel Psychology*, 75, 581–595.
- HUNTER, J. E. (1983a). *The dimensionality of the General Aptitude Test Battery (GATB) and the dominance of general factors over specific factors in the prediction of job performance*. Washington, DC: U.S. Employment Service; USES Test Research Report No. 44.
- HUNTER, J. E. (1983b). Test validation for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB). Washington, DC: U.S. Employment Service, USES Test Research Report No. 45.
- HUNTER, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, 29, 340–362.
- HUNTER, J. E., & HIRSCH, H. R. (1987). Applications of meta-analysis. In C. L. Cooper and I. T. Robertson (Eds.), *Review of industrial psychology*, Vol. II, 1987 (pp. 321–357). New York: Wiley.
- HUNTER, J. E., & HUNTER R. F. (1984). Validity and utility of alternate predictors of job performance. *Psychological Bulletin*, 96, 72–98.
- ONES, D. C., VISWESVARAN, C., & SCHMIDT, F. L. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology*, 78, 679–703.
- UNITED STATES EMPLOYMENT SERVICE (1970). *Manual for the USES General Aptitude Test Battery. Section III: Development*. Washington, DC: U.S. Department of Labor, Manpower Division.
- SCHMIDT, F. L., & HUNTER, J. E. (1992). Causal modeling of processes determining job performance. *Current Directions in Psychological Science*, 1, 89–92.

JOHN E. HUNTER

GENERAL INTELLIGENCE Psychometricians have developed a large number of standardized tests designed to measure performance over broad areas. Such tests are referred to as measures of “intelligence,” “IQ,” “general cognitive ability,” or “scholastic ability,” just to mention a few terms. However, even though many different concepts are used, the tests are usually positively correlated and are commonly regarded as measures of different aspects of general intelligence (see Carroll, 1982). Referring to this concept, S. Scarr (1989, p. 75) observed that “no concept in the history of psychology has had or continues to have as great an impact on everyday life in the Western world.” It might be added that few concepts have generated as much controversy as has the concept of general intelligence.

The fundamental question of the nature of general intelligence has received many different answers (see Sternberg & Detterman, 1986), and a wide variety of different research approaches have been employed to elucidate the nature of intelligence (Sternberg, 1990). In much research and in most applications, tests of intelligence have played a central role, which raises the question of what the range of behavior is that indicates human intelligence. Another question is whether measured intelligence should be viewed as a unitary phenomenon or as a multifaceted phenomenon. Yet another controversial issue concerns the relative weight that should be given to general intelligence and to specialized abilities, some taking the position that general intelligence is the most important dimension of individual differences and others arguing that it is of limited importance.

INTELLIGENCE AND INDIVIDUAL DIFFERENCES

The first practically useful intelligence test was developed by Alfred BINET and Théophile Simon (1905, 1908). The test consisted of many different kinds of tasks, most of which were quite complex. The first version of the scale was presented as a set of clinical procedures that did not result in a summary score, but the revision of 1908 did provide a summary score (“mental level”), which was developed into the concept of intelligence quotient (IQ) by the German psy-

chologist Louis W. Stern and was adopted by L. M. TERMAN (1916).

Measures expressing an overall level of intellectual performance have ever since been prominent in applications and measurement devices. Even in test batteries designed to measure several narrow abilities, the scores are often combined into an overall measure of general intelligence. Such complex measures of general intelligence are related to performance in many educational and occupational areas and are thus practically useful, but the usefulness of general intelligence as a theoretical concept has been questioned.

One of the most controversial questions in research on intelligence is whether individual differences in performance may be understood in terms of an underlying general ability or in terms of a collection of more or less unrelated abilities. The empirical research on this problem has typically investigated profiles of performance over different tasks, using tools such as the coefficient of correlation. Methods of multivariate analysis have also been applied to analyze patterns of intercorrelations between large sets of tasks in terms of a limited number of underlying dimensions.

The single most fundamental observation in support of the notion of general intelligence is the fact that correlations between different measures of cognitive performance are positive, even though the correlations are far from being perfect. Nonetheless, even though positive covariation is a necessary requirement for such a concept to be reasonable, there are several alternative explanations that do not involve the notion of general intelligence.

Charles SPEARMAN (1904) developed the first formalized model to explain the phenomenon of positive correlations among task performances. The model assumes that an observed score for an examinee may be accounted for in terms of a weighted sum of scores on two underlying unobservable variables: one general, which is common to all tasks (g), and one specific to each task (s). From this model, it follows that the correlation between any two variables is a function of the amount of relationship between each of the variables and the g factor. If the empirically observed correlation is close to the theoretically expected correlation, the model fits the data, but if there are differences, the model does not render a plausible simplification of the observed relations between variables.

In his empirical work Spearman used small samples of variables and subjects, and often he did find a very good fit between the observational data and the model, although deviations were noted as well. The s factors often were found to be correlated, thus giving rise to additional common factors. The model also broke down when tests that were “too similar” were included in a battery of tests, again because of a correlation between the s factors.

During the 1930s, multidimensional alternatives to Spearman’s theory appeared. L. L. THURSTONE (1938, 1947) extended Spearman’s simple unidimensional model to encompass multiple factors. In an application of multiple factor analysis, as the method was called, Thurstone (1938) designed a battery of tests in which about half a dozen primary mental abilities (PMAs) were identified: The analysis did not indicate a general factor (see PRIMARY MENTAL ABILITIES). In further studies, several of the original PMAs were decomposed into more narrow factors, and new factors were discovered when new domains were investigated. Thus, narrow factors proliferated and the general factor disappeared from the factor-analytic models of the structure of abilities.

The proliferation of narrow abilities made it necessary to bring order to the multitude of factors. One way to do this is to analyze the correlations between the factors with factor analysis to obtain higher-order factors. Such higher-order analyses yield hierarchical models, in which factors at lower levels are subsumed under factors at higher levels.

L. L. Thurstone and T. G. Thurstone (1941) analyzed the intercorrelations of the PMAs and found a general factor. A more elaborate hierarchical model (the gf - gc model), which includes two general dimensions (crystallized intelligence, gc , and fluid intelligence, gf) and several other broad abilities, has been developed by R. B. CATTELL and J. L. Horn (e.g., Cattell, 1963; Horn, 1968; Horn and Cattell, 1966) (see FLUID AND CRYSTALLIZED INTELLIGENCE, THEORY OF). A very similar hierarchical organization has been presented by J. B. Carroll (1993), on the basis of reanalyses of a very large number of matrices of correlations analyzed to indicate structure among abilities. Carroll’s three-stratum theory includes factors of three degrees of generality: narrow, broad, and general. The general factor is a single factor at the third and highest level

of the model, and this factor influences performance in each and every domain. These hierarchical models thus restore the concept of general ability and combine the perspective of those emphasizing several narrow dimensions of ability with the perspective of those emphasizing one general cognitive ability.

Another way of defining the general factor is based on principal-components analysis. In principal-components analysis (Hotelling, 1933), the observed variables are transformed into new uncorrelated (orthogonal) variables in such a way that the first component accounts for the largest proportion of the total variance, the second-largest principal component for the second-largest proportion, and so on. The components include both the common and the specific factors—that is, the total test variance. Some researchers identify the first component as representing Spearman's *g* factor of general intelligence (see, e.g., Ceci, 1990; Jensen, 1980, 1982), although, as noted, Spearman's model specified that the common factor alone indicates *g* (Horn, 1989).

Whether the general factor is defined as a higher-order factor or as a first principal component, the nature of the factor varies as a function of which particular tests are included in the matrix (e.g., Horn, 1989; Thurstone, 1947). If there are many verbal tests in one battery, for example, the first principal component is heavily weighted with verbal abilities. Common factor analysis with confirmatory tests at the higher order address this problem. The results of several such studies indicate that the *g* factor is, in the batteries analyzed, not distinguishable from the broad ability *Gf* (see FACTOR ANALYSIS). If this relation holds up in further empirical research, it provides a basis for establishing an invariant general factor, which is similarly defined as Spearman's *g* factor (Horn, 1980; Undheim, 1981).

Evidence in favor of a general dimension of cognitive ability has been obtained from studies using other methods of analysis as well. In multidimensional scaling, a space is constructed within which the tests are represented as a geometric configuration of points. Building upon techniques and concepts originally contributed by L. Guttman (1966, 1970), R. E. Snow, P. C. Kyllonen, and B. Marshalek (1984) proposed a model, the so-called radex model, which yields a two-dimensional map of tests (see FACET ANALYSIS; RADEX

THEORY). In the center of the radex map are complex tests that typically are highly related to *gf*. The radex is also divided into content areas (verbal, numerical, and figural), and within each content area, the tests are ordered according to complexity. In the numerical domain, for example, tasks requiring application of arithmetic skills are closest to the periphery; tasks requiring greater problem solving are closer to the center; and closest to the center are highly complex tests. In the other content areas as well, progressions from less complex to more complex tasks appear. Snow, Kyllonen, and Marshalek also showed that the factor loadings of tests on the general factor were almost perfectly related to the level of complexity of the tests, as defined by the multidimensional scaling. Thus, the complexity dimension of the radex model seems to be identical with the general dimension of the factor model.

Most present-day models of the structure of abilities either include a general ability or, as in Horn's theory, allow for this possibility if the data warrant it. H. Gardner (1983) has proposed a classification of abilities that does not leave room for general ability. Gardner's MULTIPLE INTELLIGENCES THEORY has primarily been developed from sources of information other than those based on psychometric analyses. On the basis of a review of literature from several areas, such as developmental psychology, neuropsychology, and cross-cultural research, Gardner has proposed a list of seven intelligences: linguistic, logical-mathematical, spatial, musical, bodily-kinesthetic, interpersonal, and intrapersonal. These concepts are descriptively similar to nine intelligences of *Gf-Gc* theory (Horn, 1989) and to dimensions within Carroll's three-stratum model. The intelligences are assumed to work in concert in the solution of particular problems, but Gardner hypothesizes that the intelligences are independent in the sense that level of performance achieved by one intelligence is not related to the level achieved by the other intelligences. Gardner explains the observed correlation between the psychometrically identified abilities as a result of the fact that tests of intelligence are paper-and-pencil exercises that rely heavily on linguistic and logical-mathematical abilities.

Gardner's rejection of general intelligence should not be interpreted too strictly, for he ascribes general capacities of pattern perception to logical-mathemati-

cal ability, which implies that it may cause performances in different domains to be correlated and thus have properties of a general dimension. Gardner's conclusion that the intelligences are independent has also been criticized for lacking a firm empirical basis (Brody, 1992, pp. 36–40).

In summary, it seems that considerable empirical evidence supports a dimension of general mental ability, which is most heavily involved in tasks that require complex processes, such as abstraction, rule inference, generalization, and transformation of content (cf. Jensen, 1992, p. 274). Again it may be noted that the empirical results are well in line with Spearman's (1923) early formulations about characteristics of optimal *g*-tasks, which should involve "eduction of rules" (rule inference) and "eduction of correlates" (rule application).

INTERPRETATIONS OF GENERAL INTELLIGENCE

Many different interpretations of general intelligence have been proposed. Spearman (1904, 1927) viewed the *g* factor as a unitary phenomenon and suggested that individual differences in general ability could be accounted for in terms of differences in "mental energy." Binet, in contrast, did not conceive of general intelligence as a unitary dimension but as a mean level determined by several capacities, such as judgment, flexibility, and goal-directedness (see Cronbach, 1990, pp. 229–230). Binet thus saw the score as representing an average of several abilities. Spearman objected to this position and argued that the conception of general mental ability as an average leads into difficult theoretical problems, primarily because the domain over which the average is to be taken is not defined.

Spearman's idea that *g* is a singular entity related to the overall energy of the mind was challenged by other researchers, in particular in this century by Godfrey THOMSON (e.g., 1916), who accounted for the *g* factor by assuming that there is a large number of bonds or units in the mind, subsets of which are sampled by different tests. The rather heated debate between Thomson and Spearman did not end in victory for one of the two interpretations (see Brody, 1992, pp. 10–

13). Both interpretations explain the existence of a general factor, but they have not been specified in ways that allow them to be distinguished empirically.

The discussion of whether general intelligence should be conceived of as a unitary or as a complex phenomenon has continued. R. E. SNOW and D. F. LOHMAN (1989) pointed out that in recent cognitive research in which processes of solving test items were studied, the latent-ability variable that accounted for the processes is not univocal; it is merely a convenient summary of the number of items correct on the tests. This position is more in line with Thomson's thinking than with Spearman's. L. G. HUMPHREYS (1985, 1989) also takes a position close to Thomson's. He interprets the general factor as a sampling (of bonds) of the individuals' entire repertoire of knowledge and skills. Others, such as H. J. EYSENCK (1988), argue that Spearman's principles of eduction of correlates and relations provide the best understanding of *g*.

Cronbach (1990) in a sense combines the unitary and the collection-of-bonds views by arguing that general intelligence may be viewed as a single index of capacity that is complexly determined.

To say that one person is "more intelligent" than another can only mean that he or she uses information more efficiently to serve his or her purposes. The efficiency of a factory is not to be located in this or that part of the operation. Rather, the purchasing division, the mechanics, the operators, the inspectors, and the shippers do their tasks with few errors and little lost time. Efficiency is a summary statement of what they accomplish as a team [p. 230].

Again it seems that dual perspectives are necessary to understand general intelligence.

Since the 1970s, experimental cognitive psychology has developed powerful theories and methods to account for how individuals receive, transform, store, and retrieve information. These theories and methods have also been applied to understanding individual differences in general intelligence. Thus, cognitive processes have been hypothesized and identified with chronometric techniques, which processes have then been related to performance on intelligence tests. It may be expected that as the insights into the mechanisms of the human mind improve, better interpreta-

tions of intelligence in information-processing terms will be achieved.

MEASURING GENERAL INTELLIGENCE

Two major approaches have been tried in attempts to develop optimal measures of general intelligence. In one approach, particular kinds of tasks with as high a relationship as possible to a general factor have been sought. Such tasks are designed to provide homogeneous measures of intelligence. The RAVEN PROGRESSIVE MATRICES is an example of such a test. It was developed by J. K. C. Raven (1938), based on a test of spatial analogies developed in Spearman's laboratory (Spearman, 1927), following Spearman's suggestion that an optimal g -test should involve "eduction of relations and correlates" (see Spearman's TWO-FACTOR THEORY): The empirical evidence indicates that the Raven test is highly related to estimates of general intelligence.

In other approaches to measuring intelligence, tests are constructed to provide mixtures of several different kinds of abilities. Such tests are called "heterogeneous tests." Ever since Binet introduced this approach, it has been by far the most commonly employed avenue to measuring general intelligence. Tests consisting of a mixture of items have been viewed with suspicion; they are so varied that they do not seem to yield an interpretable measure of any identifiable ability. However, a rationale for this kind of measure was provided by Spearman. He demonstrated that the particular content of a set of tasks is unimportant as long as the measure is highly related to g . If several different tasks are combined into one score, the importance of the factors typically diminishes, but the importance of g increases. Thus, the best way to obtain a pure measure may be to mix a large number of different tasks. It is interesting to observe that even though Spearman did not approve of Binet's theory of intelligence, he did sympathize with the idea "of throwing many miscellaneous tests into a single pool" (Spearman, 1927, p. 84) to obtain an estimate of g .

It seems somewhat paradoxical that a mixture of different tasks would provide a purer measure of general intelligence than would a homogeneous measure (cf. Humphreys, 1985), but if one uses the hierarchical

approach, this principle may be demonstrated mathematically. It should not be concluded that existing measures of general intelligence provide a pure and optimal measure of a general factor of intelligence. The reason for this is that these tests have not been constructed to be as broad as possible but to be good predictors of school achievement and other criteria. As a consequence, there is an overrepresentation of verbal tasks in such tests, so most current tests of intelligence are likely to involve quite an important dimension of verbal ability in addition to general ability.

CONCLUSION

For a long time the notion of general intelligence played an important role in practical applications, but then research decreased as more studies were directed at understanding components of intelligence. However, through increased emphasis on hierarchical models and through process-oriented research, general intelligence has received increasing attention as a central dimension of individual differences. The dimension is not quite well established empirically, but it must be emphasized that there are several competing concepts of, and theories about, general intelligence. There is evidence to support several features of those theories, but much remains to be learned before a theory acceptable to all the evidence will be available. Nevertheless, the concepts, theories, and measures of general intelligence are important both in guiding research and in practical applications.

It should be emphasized that general intelligence as a dimension of individual differences should be ascribed a proper amount of influence and that the degree of importance of this dimension should be neither overrated nor underrated. The most important consideration in this context is whether the domain of performance that we want to make statements about is broad or narrow. It is quite a general principle that performance on particular intellectual tasks is less highly related to measures of general intelligence, but measures of intellectual performance over a wider area, such as school achievement over the course of a year, typically is relatively highly related to general intelligence. The reason for this is that the relative importance of specialized abilities is greater in narrow

classes of tasks than in broad classes of tasks. Thus, even though there is virtually no situation in which the concept of general intelligence may be completely disregarded, there also are many situations in which more specialized ability concepts should be considered along with general intelligence.

BIBLIOGRAPHY

- BINET, A., & SIMON, T. (1905). Méthodes nouvelles pour le diagnostic du niveau intellectuel des anormaux (New methods for diagnosing the intellectual level of abnormals). *L'Année Psychologique*, 11, 191–336.
- BINET, A., & SIMON, T. (1908). Le développement de l'intelligence chez les enfants (The development of intelligence in children). *L'Année Psychologique*, 14, 1–94.
- BRODY, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.
- CARROLL, J. B. (1982). The measurement of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. New York: Cambridge University Press.
- CARROLL, J. B. (1993). *Human cognitive abilities; A survey of factor-analytic studies*. New York: Cambridge University Press.
- CATTELL, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1–22.
- CECI, S. J. (1990). *On intelligence . . . more or less: A bioecological treatise on intellectual development*. Englewood Cliffs, NJ: Prentice-Hall.
- CRONBACH, L. J. (1990). *Essentials of psychological testing* (5th ed.). New York: Harper & Row.
- EYSENCK, H. J. (1988). The concept of "intelligence": Useful or useless? *Intelligence*, 12(1), 1–16.
- GARDNER, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- GUSTAFSSON, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179–203.
- GUSTAFSSON, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg, *Advances in the psychology of human intelligence* (Vol. 4). Hillsdale, NJ: Erlbaum.
- GUTTMAN, L. (1966). Order analysis of correlation matrices. In R. B. Cattell (Ed.), *Handbook of multivariate experimental psychology*. Chicago: Rand McNally.
- HORN, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, 72, 242–259.
- HORN, J. L. (1980). Concepts of intellect in relation to learning and adult development. *Intelligence*, 12, 285–317.
- HORN, J. L. (1989). Models of intelligence. In R. L. Linn (Ed.), *Intelligence*. Urbana: University of Illinois Press.
- HORN, J. L., & CATTELL, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, 57, 253–270.
- HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, 24, 417–441, 498–520.
- HUMPHREYS, L. G. (1985). General intelligence. An integration of factor, test and simplex theory. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications*. New York: Wiley.
- HUMPHREYS, L. G. (1989). Intelligence: Three kinds of instability and their consequences for policy. In R. L. Linn (Ed.), *Intelligence*. Urbana: University of Illinois Press.
- HUNT, B. (1978). Varieties of cognitive power. In L. B. Resnick (Ed.), *The nature of intelligence*. Hillsdale, NJ: Erlbaum.
- JENSEN, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- JENSEN, A. R. (1982). The chronometry of intelligence. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum.
- JENSEN, A. R. (1992). Understanding *g* in terms of information processing. *Educational Psychology Review*, 4(3), 271–308.
- RAVEN, J. K. C. (1938). *Progressive matrices: Sets A, B, C, D, & E*. London: H. K. Lewis.
- SCARR, S. (1989). Protecting general intelligence: Constructs and consequences for interventions. In R. L. Linn (Ed.), *Intelligence*. Urbana: University of Illinois Press.
- SNOW, R. E., KYLLONEN, P. C., & MARSHALEK, B. (1984). The topography of ability and learning correlations. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol 2). Hillsdale, NJ: Erlbaum.
- SNOW, R. E., & LOHMAN, D. F. (1989). Implications of cognitive psychology for educational measurement. In R. Linn (Ed.), *Educational Measurement* (3rd ed.). New York: Macmillan.
- SPEARMAN, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201, 293.
- SPEARMAN, C. (1923). *The nature of 'intelligence' and the principles of cognition*. London: Macmillan.

- SPEARMAN, C. (1927). *The abilities of man*. London: Macmillan.
- STERNBERG, R. J. (1990). *Metaphors of mind: Conceptions of the nature of intelligence*. Cambridge: Cambridge University Press.
- STERNBERG, R. J., & DETTERMAN, D. K. (EDS). (1986). *What is intelligence? Contemporary viewpoints on its nature and definition*. Norwood, NJ: Ablex.
- TERMAN, L. M. (1916). *The measurement of intelligence*. Boston: Houghton-Mifflin.
- THOMSON, G. H. (1916). A hierarchy without a general factor. *British Journal of Psychology*, 8, 271–281.
- THURSTONE, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, 1.
- THURSTONE, L. L. (1947). *Multiple factor analysis*. Chicago: University of Chicago Press.
- THURSTONE, L. L., & THURSTONE, T. G. (1941). Factorial studies of intelligence. *Psychometric Monographs*, 2.
- UNDHEIM, J. O. (1981). On intelligence IV: Toward a restoration of general intelligence. *Scandinavian Journal of Psychology*, 22, 251–265.

JAN-ERIC GUSTAFSSON

GENETICS, BEHAVIOR Behavior genetics is the study of the genetic and environmental sources of individual differences in behavioral traits. An interest in the constitutional and acquired features of human conduct goes back at least to the ancient Greeks, but the immediate ancestor of modern behavior genetics was the multitalented nineteenth-century Englishman FRANCIS GALTON.

In his book *Hereditary Genius* (1869), FRANCIS GALTON explored the tendency of excellence of various kinds to run in families, a tendency he believed to be largely due to biological heredity. In other writings, Galton developed the statistical ideas of correlation and regression, which are central tools of today's behavior geneticists, and he noted the relevance of twins to the study of heredity and environment. Galton did not have at his disposal the modern distinction between monozygotic and dizygotic twins, but he did note that there was a subgroup of strikingly similar twins, and suggested that one could study the power of environment to shape behavior by starting out either with highly similar pairs to see how different they became

with experience, or with initially dissimilar pairs subjected to similar environments to see whether this made them more alike.

One of the first behavior-genetic studies using ability tests was that of E. L. THORNDIKE in 1905 (Rende, Plomin, & Vandenberg, 1990). Thorndike pursued Galton's idea by examining the resemblance of older and younger sets of twins. However, he did not believe in two distinct subvarieties of twins, and it was not until the 1920s that modern twin studies of IQ emerged. The 1920s also saw the first adoption studies of IQ (Burks, 1928; Freeman, Holzinger, & Mitchell, 1928).

A scattering of nature-nurture studies of IQ and other traits continued through the 1930s, 1940s, and 1950s in the United States, Britain, and Europe. In 1960, with the publication of the textbook *Behavior Genetics* by J. L. Fuller, a biologist, and R. F. Thompson, a psychologist, the contemporary field of behavior genetics could be considered to be established. Fuller and Thompson's book contained a chapter on intellectual abilities, surveying studies on animals and humans up to that time. More recent reviews of behavior-genetic methods and findings for intelligence and other traits may be found in texts such as Hay (1985) and Plomin, DeFries, and McClearn (1990), and in periodic chapters on the topic in the *Annual Review of Psychology*. The key journal in the field is *Behavior Genetics*.

ESTIMATING GENETIC AND ENVIRONMENTAL CONTRIBUTIONS TO IQ

Behavior genetics attempts to assess how differences in a trait in a population of individuals can be explained by differences among the genes of the individuals or among their environments. It is important to emphasize that the concern is with *differences*. For any one person, it is rarely possible to specify the relative contributions of genes and environments to the development of any characteristic, because the two have been continuously interacting in immensely complex ways since conception. D. O. HEBB (1972) suggests the analogy of trying to separate the contributions of width and length to the area of a field. It is impossible; without either, the field simply doesn't exist. Nevertheless, if Hebb's analogy is extended to a

population of fields, matters are different. It is perfectly feasible to ask to what extent the areas of a group of fields depend on differences among their widths or among their lengths. For example, if the fields are all of about the same width, their differences in area will mostly be attributable to their differences in length. If they are all of the same length, their areas will reflect their differences in width. For intermediate cases, one can assess the relative contributions of length and width to the variation in areas.

For lengths and widths, read genes and environments. By methods shortly to be described, using twins, adoptions, and other relevant family configurations, it is possible to decompose the variation of a trait (such as intelligence) in a particular population into components associated with genetic variation in that population, with variation in the environments that members of the population have encountered, and with correlation and interaction (in the statistical sense) between the two.

Any such variance decomposition refers to the particular population studied. If other populations have greater or less variation among the genes or the envi-

ronments of their members, or if the amount of correlation or interaction differs, the apportionment can come out differently. However, in practice, large population-to-population differences appear to be uncommon.

Initially, methods and results will be discussed for the trait of general intelligence, as measured by typical IQ tests. Later, the issue of possible differences for different kinds of intellectual performance, such as verbal and spatial abilities, will be considered.

IQ CORRELATIONS AMONG RELATIVES

Table 1 is a summary of resemblances in IQ among various kinds of relatives, taken from a compilation by Bouchard and McGue (1981) of studies in the literature. The values given in the table are the averages of anywhere from two to sixty-nine correlations from different studies, weighted by sample size; the total number of pairings of individuals involved is given in the rightmost column.

The equations shown in the table represent a theoretical model of the way in which these resemblances

TABLE 1
Average IQ correlations among various relatives

Equation	Average Correlation	Number of Pairings
$r_{MZA} = h^2 + d^2$.72	65
$r_{SA} = .5 h^2(1 + h^2m) + .25 d^2$.24	203
$r_{POA} = .5 h^2(1 + m)$.24	720
$r_{MZT} = h^2 + d^2 + c_T^2$.86	4,672
$r_{DZT} = .5 h^2(1 + h^2m) + .25 d^2 + c_T^2$.60	5,533
$r_{ST} = .5 h^2(1 + h^2m) + .25 d^2 + c_S^2$.47	26,473
$r_{POT} = .5 h^2(1 + m) + C_p^2$.42	8,433
$r_{ANT} = c_S^2$.29	345
$r_{AAT} = c_S^2$.34	369
$r_{PAT} = c_P^2$.19	1,491

NOTE: Average correlations shown are weighted means. MZ = monozygotic twins, A = reared apart, T = reared together, DZ = dizygotic twins, S = biological siblings, PO = parent and biological offspring, AN = siblings, one adopted and one biological, AA = siblings, both adopted, PA = parent and adopted child. h^2 = additive genes; d^2 = genetic dominance; c^2 = shared environment: T = twins, S = nontwin siblings, P = parent and child; m = spouse correlation. Data from Bouchard & McGue (1981), with corrections (Bouchard, personal communication, 1988). Based on Loehlin (1989).

might have arisen from genetic and environmental factors shared by the individuals in question. The symbols h^2 and d^2 refer to two kinds of genetic effects, so-called additive and dominance effects of genes. Additive effects are the average effects of individual genes on the trait. Dominance effects depend on combination of genes at particular chromosomal loci. A recessive gene may or may not have an effect, depending on whether a dominant version of the gene is also present. Because gene combinations are broken up each generation, dominance effects do not contribute to parent-child resemblance, although they contribute fully to the resemblance of identical twins, who share all their genes and hence any patterns among them, and to a lesser extent to the resemblance of fraternal twins and ordinary siblings, by the odds of both inheriting the same combination from their parents.

The various c^2 terms represent the contributions of shared environmental factors to resemblances in IQ. The equations allow for this occurring at three different levels: between twins, between ordinary siblings, and between parents and offspring, symbolized respectively by c_T^2 , c_S^2 , and c_P^2 .

Finally, the term m represents the correlation between spouses, which can affect sibling and parent-child resemblances. If two parents resemble each other genetically, their children will tend to be more like one another than if the parents are very different, and a mother (say) will resemble her children because of her husband's genes as well as her own. Spouse resemblance enters into equations somewhat differently depending on how it is assumed to originate. The assumption in these equations is the one most commonly made, that the resemblance arises from the trait itself rather than specifically from its genetic or environmental antecedents; that is, that spouses are correlated for intelligence because people of similar levels of intelligence tend to wind up in similar educational, occupational, and avocational settings, and hence have a tendency to meet and marry each other, and that this process is neutral with respect to the genetic or environmental origins of the trait.

Twin Studies. Most studies that have attempted to estimate genetic and environmental contributions to behavioral traits, including intelligence, have worked with one or two groups. By far the commonest procedure has been to compare the resemblance of iden-

tical and fraternal twin pairs. Identical, or monozygotic (MZ), twins come from the splitting of a single fertilized ovum, and hence are genetically identical. Fraternal, or dizygotic (DZ), twins come from a multiple ovulation in which there is separate fertilization of two ova by different sperms. Thus, genetically, fraternal twins are ordinary siblings, although they may share prenatal and postnatal environments to a greater degree than other siblings do. If the genes are important in accounting for variation on a trait, one would expect identical twins, who share all their genes, to be more similar, on the average, than fraternal twins, who share only half theirs.

The HERITABILITY of a trait may be defined as the extent to which the genes contribute to its variation. (This is sometimes called heritability in the broad sense, represented by $h^2 + d^2$ in the Table 1 equations. Heritability in the narrow sense, h^2 , means just the additive genetic contribution.) A number of formulas for estimating the heritability of a trait from a twin study have been suggested. The simplest is just to double the difference between the identical and fraternal intraclass correlations:

$$\text{heritability} = 2(r_{MZ} - r_{DZ}).$$

An examination of the expression for r_{MZT} and r_{DZT} reveals that the heritability formula above will be straightforward only if m and d are both zero. For a number of psychological traits, these assumptions may be plausible, but for IQ, there is known to be appreciable positive correlation between spouses, and there is evidence from studies of inbreeding (discussed below) that genetic dominance makes a contribution. These two factors tend to act in opposite directions on the formula, however, so if both are present, they will tend to offset each other. Thus, the estimate from the above expression of a heritability for IQ of $2(.86 - .60) = .52$ may not be grossly in error. Other estimates will be presented below, but for the moment this suggests that something like half the variance of IQ is due to the genes, at least in the U.S., British, and northern European populations, from which these data predominantly come.

Studies of Adoptive Families. After twin studies, the next most popular way of estimating heritability is from studies comparing parent-child correlations and sibling correlations in adoptive and

biologically related families. The resemblance among adoptive family members will be due to the environment they share, whereas the resemblance between biologically related family members will involve genetic as well as environmental resemblance. Again, a rough formula is to double the difference to estimate heritability:

$$\text{heritability} = 2(r_{\text{ST}} - r_{\text{UT}}),$$

or

$$\text{heritability} = 2(r_{\text{POT}} - r_{\text{PAT}}),$$

where r_{UT} refers to unrelated children reared together as siblings (a combination of groups r_{AAT} and r_{ANT} in Table 1). Again, examination of the equations in the table suggests that these expressions require the absence of spouse correlations and genetic dominance if they are to be accurate. Their derivation assumes also the absence of selective placement by the adoption agency that is it assumes that there was no matching of the child to the adopting family in terms of genetic potential. A moderate degree of such selective placement is often reported for IQ in such studies (e.g., Horn, Loehlin, & Willerman, 1979; Scarr & Weinberg, 1978) but is not invariably found (e.g., Plomin, DeFries, & Fulker, 1988). Again, therefore, the application of these equations will provide only rough estimates of heritability: in the present case,

$$2(.47 - .316) = .31$$

and

$$2(.42 - .19) = .46,$$

respectively. Both suggest an appreciable contribution of the genes to IQ variation, although they are a little lower than the heritability estimate yielded by the twin studies.

Studies of Identical Twins Reared Apart. A third, very simple estimate of heritability may be obtained from the study of identical twins reared apart. This is just the correlation itself:

$$\text{heritability} = r_{\text{MZA}}.$$

For the Table 1 data, this yields an estimate of 72 percent for the heritability of IQ, a figure somewhat higher than those yielded by the twin studies and the adoptive-biological comparisons. This estimate is based on considerably less data than are the others—a total

of sixty-five pairs from three studies—because identical twins who have been reared apart are rare and difficult to locate. However, an additional sample of forty-eight pairs of identical twins reared apart has since been studied (Bouchard, Lykken, McGue, Segal, & Tellegen, 1990). The IQ correlation, and hence the heritability estimate, for this additional group is .69, close to the earlier figure, thus providing some grounds for believing that the higher estimate from this source of evidence is not just a freak of sampling, but reflects the presence of nonadditive genetic effects or other factors.

Combined Model Fitting. In recent years, with the advent of structural modeling methods and associated computer programs, it has become increasingly popular to deal with sets of equations, such as those in Table 1, all at once, yielding a weighted best fit to the data and a chi-square test of the goodness of that fit, along with optimum estimates of the individual parameters. Such methods also permit straightforward comparisons of models embodying more or fewer restrictive assumptions, with simple chi-square tests of differences between them (for details, see Boomsma, Martin, and Neale, 1989, or Loehlin, 1992).

By way of illustration, consider an analysis of the Table 1 data, testing a suggestion by Robert Plomin (Plomin & Loehlin, 1989) that heritability estimates made by so-called direct methods (involving just a single correlation, as for identical twins reared apart) seem to yield higher heritability estimates than those made by indirect methods (involving a comparison of two correlations, as for identical and fraternal twins, or adoptive and biological siblings). In Table 1, direct estimates are obtainable from the first three equations, involving genetically related individuals reared apart, and indirect estimates via various pairings of the remaining equations. The Table 1 equations were fitted to the average correlations, with m , the spouse correlation, set at .33, also obtained from the Bouchard and McGue (1981) review. Fitting the full set of equations as given yielded a significantly poor fit to the data ($p < .05$). Allowing for different h^2 's for the first three and for the remaining equations gave a statistically significant improvement in fit, and the fit of the model to the data was now acceptable. This solution yielded h^2 's of 41 percent and 30 percent from the direct and indirect equations, a d^2 of 17 percent, and c^2 's of 39

percent, 27 percent, and 22 percent for the shared environments of twins, siblings, and parents and their children, respectively. Thus, overall, the broad heritability of IQ ($h^2 + d^2$) was estimated at 58 percent via the direct methods and 47 percent via the indirect ones.

What does the difference in heritabilities estimated by the two methods mean? It probably does not have anything to do with the functioning of the genes as such, which is presumably not affected by how they will eventually be studied. It may, however, provide some information about family environments. The three direct equations in Table 1 are all based on the correlations between pairs of individuals living in *different* families. The seven indirect equations are all based on correlations between pairs of individuals living in the *same* family. This suggests some systematic process within families leading to an exaggeration of differences among more similar pairs. Further research will be required to establish the details.

INBREEDING AND GENETIC DOMINANCE

If a trait is subjected to natural selection, it is expected that its additive genetic variation will be depleted more rapidly than will variation due to genetic effects such as dominance and epistasis (the latter term refers to configurational genetic effects involving more than one chromosomal locus). Selection tends to act more slowly on nonadditive genetic effects, because individual genes are less directly associated with their consequences. Thus a trait that has a relatively large ratio of nonadditive to additive genetic variance may be a trait that has been under strong natural selection. Substantial differences in brain size and cultural complexity between modern humans and their remote ancestors suggest that intelligence is a trait that has undergone such selection. Model fitting along the lines already described suggests that the genetic influences on IQ involve dominance. Another approach is to examine the effect of inbreeding. Matings between related individuals increase the likelihood of matching up recessive genes; thus the presence of dominance can be detected by *inbreeding depression* in the offspring of the matings of relatives. A number of studies have been done in which the IQs of children of the mar-

riages of relatives have been compared to those of children of the marriages of unrelated persons. Typically, some degree of inbreeding depression is found for intelligence. An example is a study by Agrawal, Sinha, and Jensen (1984), who studied the children of marriages between cousins among Indian Muslims and found an average depression on the Raven Progressive Matrices of about half a standard deviation relative to a comparison group of nonrelative marriages; their article contains references to eleven earlier studies in various countries reporting similar results.

AGE TRENDS IN IQ HERITABILITY

Twin Studies. A recent meta-analysis of twin studies by McCartney, Harris, and Bernieri (1990) reviewed sixteen studies between 1967 and 1985 that reported IQ correlations between twins and also reported the ages of the twins studied. Fifteen of these studies included both identical and fraternal twins; one included only identical twins. The authors correlated the reported twin correlations in these studies with the average age of the twins. The total range of ages was from 1 to 59 years, but the bulk of the studies were of children: the median age was just under 8 years. The obtained correlation across studies between twin resemblance and average age was .15 for identical twins and $-.25$ for fraternal twins. This means that estimates of the effects of genes on IQ based on the difference between correlations between identical twins and correlations between fraternal twins would tend to go up with age, and estimates of the effects of shared environment on IQ would tend to go down.

A similar finding emerges from a major longitudinal study of twins, the Louisville Twin Study (Wilson, 1983). Twins were recruited by means of birth records and measured repeatedly at ages from 3 months to 15 years. At first measurement, IQ correlations for identical twins were very similar to those for fraternal twins (.66 and .67, respectively), suggesting very little contribution of the genes in accounting for individual differences in performance on the infant test used. Over the first five years of life, the IQ correlation for fraternal twins remained at about the same level, but the IQ correlation for identical twins climbed to about .85, with the difference indicating a moderate degree of genetic influence (38 percent by the simple formula

given earlier). After age 5, the correlation for identical twins increased only slightly, to .88 at age 15, while the correlation for fraternal twins began to show an irregular decline, dropping to .54 for the 15-year-olds. The heritability, estimated from twice the difference between the correlations, had increased to 68 percent. Overall, the directions of change correspond to those found by McCartney, Harris, and Bernieri (1990) in their compilation of cross-sectional twin studies—an increase with age for the correlation between identical twins, and a decrease for the correlation between fraternal twins. These changes suggest both an increase in heritability with age and a decrease in the extent to which shared environmental factors make family members alike. The latter is also seen quite clearly in studies of adoptive families.

Studies of Adoptive Families. In 126 adoptive families in Texas, genetically unrelated children reared together as siblings had an IQ correlation of .17 at the time of initial testing (their ages varied, averaging about 8 years); at retesting, ten years later, these same pairs, now late adolescents and young adults, had an IQ correlation of zero (Loehlin, Horn, & Willerman, 1989). Having lived together in the same families almost since birth had created *no* enduring resemblance in IQ, although it had had some effect at early ages. Adoptive parent-child correlations also showed a drop-off with age, although not quite so dramatically. In the original testing the parent-child correlations were .19 for fathers and .13 for mothers. Ten years later, they were .10 for fathers and .05 for mothers.

A comparison of two studies of adoptive families in Minnesota (Scarr & Weinberg, 1977, 1978, 1983) shows differences of the same kind. In the first, a study of interracial adoptions, there were IQ correlations of .27 and .21 between adoptive parents and their unrelated adopted children at an average age of 7 years, and a correlation of .44 between adoptive siblings. In the second study, of a different group of ordinary adoptions, the IQ correlations between parents and adoptive children aged 16 to 22 were .16 and .09 and the correlation between adoptive siblings was $-.03$.

Model Fitting. There has been considerable interest in fitting gene-environment models to longitudinal behavior-genetic data on IQ (e.g., Vogler, 1992). These modeling efforts have differed in a number of ways, and their results have not always been in agree-

ment, but some of them have suggested that shared environmental effects tend to be substantial early, whereas new genetic contributions continue to enter the picture during the course of development, with effects that persist over time.

GENOTYPE-ENVIRONMENT (GE) CORRELATION AND INTERACTION

GE Correlation. Genetic and environmental influences on a trait may be correlated (GE correlation), or they may interact (GE interaction). The term *GE correlation* refers to a tendency for genetic and environmental influences on a trait to be associated. For example, parents with high IQs may provide their children with genes conducive to the development of high IQs, but may also provide them with environments containing above-average levels of intellectual stimulation. This is referred to as a *passive* GE correlation, because it does not depend on the behavior of the child in question (Plomin, DeFries, & Loehlin, 1977). Two other forms of GE correlation are *reactive* (or *evocative*) and *active*. Reactive GE correlation involves the reactions of other people to the child's gene-influenced behavior (e.g., a bright child may receive special encouragement from teachers). Active GE correlation involves an active seeking out or creation of a relevant environment by the child. Bright children may elect to spend time reading books, and hence may stimulate their further intellectual development. A central role of GE correlation in development has been emphasized by Scarr and McCartney (1983). However, some theorists prefer to treat active and reactive GE correlation simply as part of the genetic variance, since they represent ways in which genes are expressed.

Passive GE correlation can be estimated by comparing adoptive and biologically related families. In the former, the genes of children are not contributed by the parents who rear them; in the latter, they are. This will have effects on both the variance of the children's IQs and on the relative sizes of the parent-child correlations. A review of existing adoption studies suggested that a moderate passive GE correlation exists for IQ, on the order of .20 or .30 (Loehlin & DeFries, 1987). Active and reactive GE correlations have been less studied, although adoption provides possibilities here as well. An attempt to locate active and reactive

GE correlations in a Colorado adoption study involving young children did not find much in the intellectual domain, although there was a hint of some correlations of this sort for personality (Plomin, DeFries, & Fulker, 1988).

GE Interaction. By *GE interaction* is meant not the interaction of genes and environment in the development of a particular individual, but rather the interaction of genotypes and environments in a statistical sense across a population of individuals, such that particular combinations of genotypes and environments have effects unpredictable from the average effects of the two. Given that the average effects, as manifested in h^2 s, d^2 s, and c^2 s, plus errors of measurement, seem to account for much of the variance of IQ, GE interaction probably is not a quantitatively large component of individual differences in this domain. General support for this notion was provided in analyses of earlier IQ studies by Jinks and Fulker (1970). No significant effects of GE interaction on IQ were detected in a study of adoptive families in Colorado (Plomin, DeFries, & Fulker, 1988).

SEPARATE ABILITIES

The majority of behavior-genetic studies of ability have focused on general intelligence, usually defined as IQ, but there have also been studies examining genetic

and environmental influence on verbal, spatial, and other special abilities.

Table 2 presents some average correlations from a review of twin studies up to about 1972 by Nichols (1978). These correlations tend to run a little lower than correlations for IQ—this same review yielded average correlations of .82 and .59, respectively, for measures of general intelligence. The lower correlations probably reflect in part the lower reliabilities of the often rather brief measures of special abilities used in these studies. Also, many of the twin studies of special abilities involved somewhat older children; consequently, the lower correlations may partly reflect less shared environment. However, except possibly for divergent thinking, verbal fluency, and memory, the differences between the correlations of identical twins and those of fraternal twins, which provide the basis for estimating the effects of the genes, are not markedly smaller for special abilities than for general intelligence.

A meta-analysis by McCartney, Harris, and Bernieri (1990) identified twenty-nine twin studies between 1967 and 1985 that reported separate measures of verbal and performance IQ. For verbal IQ, the average correlations were .76 for identical twins, and .48 for fraternal twins; for performance IQ, they were .70 and .48, respectively. Again, these are both lower than the correlations for overall IQ from the Bouchard sum-

TABLE 2
Average correlations between identical and fraternal twins in studies of special abilities

Ability	Number of Correlations	Average Correlation		
		Between Identical Twins	Between Fraternal Twins	Identical-Fraternal Difference in Correlation
Reasoning	16	.74	.50	.24
Spatial visualization	31	.65	.41	.23
Clerical speed and accuracy	15	.70	.47	.22
Verbal comprehension	27	.78	.59	.19
Numerical reasoning and mathematics	27	.78	.59	.19
Memory	16	.52	.36	.16
Verbal fluency	12	.67	.52	.15
Divergent thinking	10	.61	.50	.11

SOURCE: Nichols, 1978.

mary (.86 and .60), the Nichols summary (.82 and .59), and the McCartney review itself (.81 and .61). Again, it is likely that the verbal and performance IQs represent samples averaging slightly older than those for total IQ, because they are mostly derived from Wechsler tests, thus excluding single-IQ instruments, such as the Stanford-Binet, which are often used with young children. Note that the differences between the correlations for identical twins and those for fraternal twins (.28 for verbal IQ and .23 for performance IQ) are as large as those typically found for overall IQ, suggesting as much genetic influence on verbal and performance IQ as on total IQ.

Special abilities tend to be correlated with one another, and hence with a general intellectual factor (*g*). Does their heritability depend solely on this shared *g* factor, or are their specific aspects also heritable? This question can be addressed through model fitting. On the whole, the evidence suggests at least some specificity in the genetic contributions (McGue & Bouchard, 1989). For example, a study based on a Colorado twin sample found three genetic factors underlying the intercorrelations among the subtests of the Wechsler Intelligence Scale for Children-Revised, whereas the shared environmental component of the covariance could be accounted for by a single factor (LaBuda, DeFries, & Fulker, 1987).

How fine a breakdown of abilities is desirable for behavior-genetic analysis? Do verbal and performance IQ or broad special abilities represent a sufficient level of detail? The question remains open; however, the large Hawaii Family Study of Cognition (DeFries et al., 1979) found considerable variation in degree of family resemblance for different tests purporting to measure the same ability. This suggests that a still more differentiated analysis might be worth pursuing (Plomin, 1988).

CONCLUSION

Evidence from twin studies, studies of adoptive families, and studies of inbreeding enables behavior geneticists to estimate the relative roles of the genes and the environment in contributing to individual differences in intellectual abilities in contemporary Western populations. Both genes and environment appear to

play a substantial role in accounting for individual variation; however, the enduring effects of that part of the environment that is shared with other family members appear to decrease with age, perhaps to quite low levels. The presence of an appreciable nonadditive component in the genetic variation is consistent with general intelligence's having been positively selected for during human evolution.

(See also: NATURE, NURTURE, AND DEVELOPMENT; TWIN STUDIES OF INTELLIGENCE.)

BIBLIOGRAPHY

- AGRAWAL, N., SINHA, S. N., & JENSEN, A. R. (1984). Effects of inbreeding on Raven matrices. *Behavior Genetics, 14*, 579-585.
- BOOMSMA, D. I. MARTIN, N. G., & NEALE, M. C. (Eds.). (1989). Genetic analysis of twin and family data: Structural modeling using LISREL. *Behavior Genetics, 19*, 3-161.
- BOUCHARD, T. J., JR., & MCGUE, M. (1981). Familial studies of intelligence: A review. *Science, 212*, 1055-1059.
- BOUCHARD, T. J., JR., LYKKEN, D. T., MCGUE, M., SEGAL, N. L., & TELLEGEN, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science, 250*, 223-228.
- BURKS, B. S. (1928). The relative influence of nature and nurture upon mental development. *The Twenty-Seventh Yearbook of the National Society for the Study of Education. Part I.* 219-316.
- DEFRIES, J. C., JOHNSON, R. C., KUSE, A. R., MCCLEARN, G. E., POLOVINA, J., VANDENBERG, S. G., & WILSON, J. R. (1979). Familial resemblance for specific cognitive abilities. *Behavior Genetics, 9*, 23-43.
- FREEMAN, F. N., HOLZINGER, K. J., & MITCHELL, B. C. (1928). The influence of the environment on the intelligence, school achievement, and conduct of foster children. *The Twenty-Seventh Yearbook of the National Society for the Study of Education. Part I.* 103-217.
- FULLER, J. L., & THOMPSON, W. R. (1960). *Behavior genetics*. New York: Wiley.
- GALTON, F. (1869). *Hereditary genius*. London: Macmillan.
- HAY, D. A. (1985). *Essentials of behavior genetics*. Melbourne: Blackwell.

- HEBB, D. O. (1972). *Textbook of psychology* (3rd ed.). Philadelphia: Saunders.
- HORN, J. M., LOEHLIN, J. C., & WILLERMAN, L. (1979). Intellectual resemblance among adoptive and biological relatives: The Texas Adoption Project. *Behavior Genetics*, *9*, 177–207.
- JINKS, J. L., & FULKER, D. W. (1970). Comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychological Bulletin*, *73*, 311–349.
- LABUDA, M. C., DEFRIES, J. C., & FULKER, D. W. (1987). Genetic and environmental covariance structures among WISC-R subtests: A twin study. *Intelligence*, *11*, 233–244.
- LOEHLIN, J. C. (1989). Partitioning environmental and genetic contributions to behavioral development. *American Psychologist*, *44*, 1285–1292.
- LOEHLIN, J. C. (1992). *Latent variable models: An introduction to factor, path, and structural analysis* (2nd ed.). Hillsdale, NJ: Erlbaum.
- LOEHLIN, J. C., & DEFRIES, J. C. (1987). Genotype-environment correlation and IQ. *Behavior Genetics*, *17*, 263–277.
- LOEHLIN, J. C., HORN, J. M., & WILLERMAN, L. (1989). Modeling IQ change: Evidence from the Texas Adoption Project. *Child Development*, *60*, 993–1004.
- MCCARTNEY, K., HARRIS, M. J., & BERNIERI, F. (1990). Growing up and growing apart: A developmental meta-analysis of twin studies. *Psychological Bulletin*, *107*, 226–237.
- MCGUE, M., & BOUCHARD, T. J., JR. (1989). Genetic and environmental determinants of information processing and special mental abilities: A twin analysis. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Vol. 5 (pp. 7–45). Hillsdale, NJ: Erlbaum.
- NICHOLS, R. C. (1978). Twin studies of ability, personality and interests. *Homo*, *29*, 158–173.
- PLOMIN, R. (1988). The nature and nurture of cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence*. Vol. 4 (pp. 1–33). Hillsdale, NJ: Erlbaum.
- PLOMIN, R., DEFRIES, J. C., & FULKER, D. W. (1988). *Nature and nurture during infancy and early childhood*. Cambridge: Cambridge University Press.
- PLOMIN, R., DEFRIES, J. C., & LOEHLIN, J. C. (1977). Genotype-environment interaction and correlation in the analysis of human behavior. *Psychological Bulletin*, *84*, 309–322.
- PLOMIN, R., & LOEHLIN, J. C. (1989). Direct and indirect IQ heritability estimates: A puzzle. *Behavior Genetics*, *19*, 331–342.
- PLOMIN, R., DEFRIES, J. C., & MCCLEARN, G. E. (1990). *Behavioral genetics: A primer* (2nd ed.). New York: Freeman.
- RENDE, R. D., PLOMIN, R., & VANDENBERG, S. G. (1990). Who discovered the twin method? *Behavior Genetics*, *20*, 277–285.
- SCARR, S., & MCCARTNEY, K. (1983). How people make their own environments: A theory of genotype → environment effects. *Child Development*, *54*, 424–435.
- SCARR, S., & WEINBERG, R. A. (1977). Intellectual similarities within families of both adopted and biological children. *Intelligence*, *1*, 170–191.
- SCARR, S., & WEINBERG, R. A. (1978). The influence of “family background” on intellectual attainment. *American Sociological Review*, *43*, 674–692.
- SCARR, S., & WEINBERG, R. A. (1983). The Minnesota adoption studies: Genetic differences and malleability. *Child Development*, *54*, 260–267.
- VOGLER, G. P. (Ed.) (1992). Human developmental behavior genetics. *Behavior Genetics*, *22*, 189–264.
- WILSON, R. S. (1983). The Louisville Twin Study: Developmental synchronies in behavior. *Child Development*, *54*, 298–316.

JOHN C. LOEHLIN

GENIUS By calling someone a genius we draw attention to that person’s prodigious creative powers and acknowledge achievements in art, original thought, invention, or discovery that have been immensely valuable and influential and are beyond the reach of ordinary individuals. A genius is someone who has arrived at insights or ideas that are extraordinarily potent and original or who has created literary, artistic, or scientific masterpieces of exceptional power and inventiveness.

THE DEFINITION OF GENIUS

The modern definition of the word *genius* given above was unknown prior to the eighteenth century. For the Romans, the genius was a kind of tutelary household spirit (the Latin word *gens* means family) that was embodied in individuals and groups of people and that accompanied a person throughout life. This

notion of genius as a kind of personal spirit that attended and protected an individual remained dominant throughout the middle ages. But by the early eighteenth century, meanings of the word that are closer to modern definitions began to be encountered, as in Alexander Pope's phrase "true genius is but rare." These draw not only on the Latin word *genius* but also on the Latin *ingenium*, which refers to innate ability or natural disposition. Yet the idea that genius can be attributed to supernatural powers has been slow to fade. That early thinkers such as Plato regarded poetic genius as being related to madness or, in Aristotle's view, to melancholia may be related to an assumption that the same supernatural powers that give people wisdom and inspire them with creative ideas may also drive them out of their minds (Ochse, 1990). Suffice it to say that the majority of geniuses have not been insane, although a number have been manic-depressive.

The difficulty of arriving at a clear and simple definition of genius is an inevitable consequence of the fact that the word is often introduced to encompass phenomena that elude description in straightforward and unambiguous terms. For some writers genius explicitly means that which cannot be defined, so that when we account for the powers of a Shakespeare or a Beethoven by saying that they were possible "because he was a genius," we are in effect admitting that we do not understand the real causes (Murray, 1989). There is a widely held view that phenomena to which the term *genius* is attached are sufficiently rare and mysterious to resist being defined in terms that account for less extraordinary achievements. In addition, because of the tendency to invoke the concept of genius whenever the terms used to explain ordinary accomplishments prove clearly inadequate to account for outstanding feats, it is hardly surprising that questions like "What is genius?" and "How do we account for it?" have received vague and unsatisfactory answers.

There is no single measure or objective index of genius. One reason for this is that genius can take many different forms. A more fundamental reason is that when someone is described as being a genius, the choice of that word implies that the person is being not so much described as chosen to have a kind of accolade bestowed upon them, in the same way that a person is selected to be given, say, an Oscar or a Nobel

prize. We would not seek to provide an objective description or measure of the characteristics of an Oscar winner, except in the most imprecise and general terms (e.g., "a person who is likely to be a very good actor"), because the winning of the Oscar comes about not by a process of describing that person directly but by making judgments that are largely a reflection of the impact of that person's accomplishments on others. Similarly, when someone is called a genius, that person is not being measured or described in any objective sense. To say that someone is a genius is largely an acknowledgment that what that person has done is greatly valued and appreciated by others. Insofar as such a statement also provides any kind of rating of the person, it is a description that merely classifies the individual as belonging to the (very small) class of people who are capable of the achievements that bring about the circumstances in which the accolade of genius is bestowed.

It follows that whether a person is regarded as being a genius largely depends upon factors that are outside the individual's control. It is certainly safe to assume that anyone whose accomplishments are sufficiently acknowledged for that person to merit the label "genius" is an individual who has achieved something exceptional. Nevertheless, there have been numerous people who are indisputably exceptionally creative, clever, intelligent, talented, or artistic without having been regarded as geniuses. The difference between being outstandingly able and being considered a genius is akin to the distinction between being one of the numerous soldiers who act bravely in a battle and one of the few who are awarded a medal for heroism. Undoubtedly one has to act bravely in order to get the medal, but it is also necessary to be fortunate enough to be brave at the right place at the right time, so that one is noticed by someone who is in a position to report the courageous action to the appropriate authorities. With genius, similarly, not only is it necessary to make the kind of contribution that can be achieved only by someone who possesses outstanding abilities, but it is also essential for the contribution to be widely recognized and highly valued by other people. Whether or not that happens may depend on any of a number of factors, such as good fortune, whether the contribution is appreciated by influential people, and whether it is in tune with the artistic or scientific

climate that forms the spirit of the times. Creative achievements that are too far ahead of their time may be ignored and, if not preserved, never properly recognized.

We are most likely to recognize genius in an individual when the particular exceptional qualities the person possesses happen to be well suited to producing advances or new insights at the time when the individual is working. For example, Albert Einstein possessed a combination of striking qualities that made him extremely well equipped for making advances at a particular time in the history of science. The happy fit that existed between the abilities that Einstein mastered and the qualities that were needed to make scientific progress in his lifetime might not have existed had Einstein been born thirty years earlier or thirty years later.

In some instances an individual's contemporaries fail to recognize qualities or achievements that lead to later generations hailing the individual as a genius. Similarly, the work of the composer Antonio Vivaldi received little attention until the twentieth century, although J. S. Bach drew upon it. Bach himself was regarded more highly as an organist than as a composer in his own lifetime. Sandro Botticelli was thought to have been a second-rank painter until the nineteenth century, and during Rembrandt's lifetime he was regarded as a less talented artist than his now-forgotten contemporary Jan Lievens. It may be that a future century's list of twentieth-century geniuses will include names that are presently unknown or not regarded highly.

Sometimes the influence of a perceptive critic such as John Ruskin can persuade people to reassess the contributions of a previous generation and discern previously unrecognized contributions of genius. Doubtless there have been a substantial number of exceptional people who have produced achievements of great potential value that have gone forever unnoticed or unrecognized for their true value. For example, in the twentieth century there is the case of the Indian mathematician Srinivasa Ramanujan, a man possessed of remarkable abilities whose background and education prevented him from being in a position to have the impact on mathematics that he might well have made had he been raised in a prosperous North American or European family.

Although any person who is widely regarded as being a genius will almost certainly be an individual possessing exceptional qualities, there may be a very small number of instances in which a highly regarded achievement has been attained partly by accident—just as the occasional medal for heroism has been won by a soldier stumbling by mistake into performing an action judged by others as being heroic. It is conceivable that something of this kind might have happened in the case of Gregor Mendel. The powerful implications of his experiments did not become apparent until around 1900, after his death, when the theoretical framework developed by William Bateson and others gave Mendel's work an importance it had not previously possessed, probably not even to Mendel himself. As Mihaly Csikszentmihalyi (1988) implies, it is conceivable that the genius of Mendel's work lies at least to some extent in the use to which it was put by other scientists after his death, rather than in its creator's mind.

FAMILY BACKGROUNDS AND THE ORIGINS OF GENIUS

What are the human attributes that form the origins of genius, enabling an individual to reach levels of achievement beyond the reach of other people? As we have seen, no particular abilities, however exceptional, are sufficient to guarantee that someone will be regarded as a genius, because the circumstances that make this possible are largely outside the control of any person, whatever their talents and capabilities. Nevertheless, it is possible to indicate some of the qualities that are necessary, although not sufficient on their own, if someone is to become a genius.

Francis GALTON wrote an influential nineteenth-century book entitled *Hereditary Genius*. Galton took it as a certain fact that genetic factors largely account for the manifestation of genius in a few individuals. He had noticed that genius and eminence often ran in families and that eminent people have a higher probability than others of having eminent parents or other relatives. Galton himself was a close relative of Charles Darwin, who was himself related to the Huxleys, a family who have given birth over several generations to a number of eminent scientists and other distinguished individuals, including Aldous Huxley, the au-

thor. Galton admitted that qualities such as zeal and the tendency to work hard and persevere were important, in addition to inborn intellectual competence as such, but he felt that these, too, were largely inherited.

The fact that late-twentieth-century investigators place more emphasis than Galton did on the role of a person's early background and childhood experiences, factors that clearly go some ways toward accounting for the family patterns Galton identified, does not imply a denial of the importance of inheritance. It is now widely acknowledged, however, that the roles of genetic and environmental factors are not simply additive or even multiplicative. Questions about the precise manner in which genetic differences between people actually contribute to the likelihood of genius being attainable, and about the extent to which they do so, cannot be fully answered on the basis of currently available knowledge.

Investigations of the backgrounds of Nobel prize-winners have revealed that a high proportion of distinguished scientists have come from professional homes and that their fathers have often been involved in scientific or teaching activities. The home backgrounds of Nobel scientists have tended to be stable and secure. Among Nobel prizewinners in literature, however, a different pattern of typical home backgrounds is encountered, with relatively unstable family circumstances, often resulting from the death of a parent, being much more frequent than in the case of the scientists. Nobel scientists have disproportionately often come from large metropolitan areas and from Jewish family backgrounds (Berry, 1990).

It is important to exert some caution when considering the possible causes and origins of genius. When confronted with the statement that genius is x percent hard work and y percent "inspiration," it is tempting to assume that the inspiration component identifies the crucial attribute of a person of genius. On reading about the factor that singles out an unquestioned genius like Wolfgang Amadeus Mozart from other accomplished musicians, it is similarly easy to assume that the existence of a "spark of genius" in the former but not the latter is the essential element. What is actually achieved by invoking these terms, however, is in fact merely a description of outcomes and performances, a rephrasing of someone's achievement, rather

than the identification of a genuine reason for it or an explanation of the underlying qualities that form the causes of that performance. Appearances notwithstanding, the only genuine grounds for saying that someone exhibits a spark of genius reside in what the latter actually achieves: The spark of genius cannot simultaneously serve as a description of someone's exceptional capability and as an explanation that identifies the origin of that capability.

For similar reasons, it is necessary to be wary of other statements about outstandingly creative people that may at first appear to point to the causes of their being geniuses. For example, it is sometimes stated that, in people of genius, creative work is habitually fluent and effortless, or occurs outside consciousness, even while the individual is asleep. As statements describing artistic creation, these are largely wrong, although it does seem to be the case that some of the ruminations that have contributed to scientific and literary discoveries have occurred at times when the individual concerned was not giving full conscious attention to the work involved. But "effortless creation" only takes place in a mind that has been prepared by many hours of hard deliberate effort: Creative achievements are the products of hard and dedicated work. Similarly, sudden insights and mental leaps certainly do occur, but only when the ground has been well prepared by persistent and concentrated labor. If the ability to create fluently and effortlessly does sometimes appear to distinguish the person of genius from the merely competent, that ability is itself an outcome of the various causes that lead to one person but not another becoming a genius. It is not a cause or an explanation of the difference between them.

There are no shortcuts to genius. Reaching the highest levels of ability in any sphere of excellence demands, among other things, many thousands of hours of hard study and practice. For instance, to arrive at a high degree of competence as a musical instrumentalist may require as much as 20,000 or 30,000 hours of preparation, or around eight hours per day for ten years. Similarly, large amounts of time are needed to reach the highest levels of ability in other areas of achievement, such as chess (Howe, 1990; Simon & Chase, 1973). Chess masters have to be able to remember around 50,000 separate chess positions, and their ability to do so becomes possible only as an out-

come of a vast array of chess knowledge having been acquired in the course of years of training. In the case of the skill of composing music, it has been established that no musical works that are now regarded as being of major importance have been created prior to the composer's having devoted around ten years to concentrated training in composing, typically following years devoted to the acquisition of basic musical abilities (Hayes, 1981). Even in the case of Mozart, who is often believed to have produced great masterworks at the age of four or five, none of the compositions by him that are now regarded as major works were actually written earlier than the twelfth year of his musical career.

The fact that genius takes many different forms contributes to the difficulty of making a clear statement about its origins. It is true that the early formative years of geniuses almost always seem to have been filled with events that in one way or another provided a rich source of experiences that the individual could draw upon in later life. In some cases, the qualities that eventually led to someone being regarded as a genius were apparent during childhood, at which time the individual was described as being a prodigy. By no means all geniuses have been child prodigies, however, and in some cases, such as that of Darwin, the individual has given few if any indications during childhood of being potentially exceptional.

PERSONAL CHARACTERISTICS OF GENIUSES

When the search for causes prompts us to look for personal attributes and characteristics that are held by all or most of those individuals who have been regarded as geniuses, it is easier to identify such shared attributes by examining the personalities and motivations of geniuses than by attending to strictly cognitive or intellectual qualities that they might be expected to share. It would be quite wrong to assume that the distinguishing attributes of people who produce outstanding achievements are exclusively intellectual ones. Of course, qualities such as intelligence are necessary and important, but knowledge of individuals' measured intelligence makes only a modest contribution to the ability to estimate their likelihood of producing creations of genius. For example, among the bright

children identified in Lewis TERMAN's large-scale study of intelligent Californians (somewhat misleadingly entitled *Genetic Studies of Genius*), those individuals with the most exceptional IQ scores did not produce more highly creative or scholarly achievements than those whose test scores were high but not as extreme.

Geniuses tend to be intensely curious and industrious, dogged, determined, and single-minded. They have a clear sense of direction and, often, a phenomenal ability to persevere in their efforts to reach an eventual goal. As Alexander Pope observed, self-confidence and optimism are two essential ingredients of great undertakings. When Isaac Newton was asked how he had succeeded in discovering the law of universal gravitation, he said that he did so by thinking about it continually. He became totally absorbed for long periods in the problems that interested him, neglecting his meals, forgetting to sleep, and working so hard that one contemporary feared he would kill himself with study. Marie Curie toiled for years at hard physical labor to achieve her goal of producing minute amounts of radium. Einstein would concentrate totally on the task he was attempting. Like Newton, he was often entirely oblivious of social niceties and the routines of everyday reality. According to one anecdote, when asked why he maintained the curious habit of using ordinary soap for shaving cream, Einstein said that having two separate kinds of soap would be too complicated for him. Darwin attributed his success to determination, a capacity for careful observation, and sheer industry. He was convinced that he was in no respect particularly clever or quick. Mozart, even as a child, was able to concentrate totally on what he was doing. Once he gave his attention to music he was unaware of anything else. Since creative achievements often demand many hours of intensive solitary effort, it is hardly surprising that geniuses are sometimes withdrawn, introspective, and introverted individuals.

Every individual genius is unique. It is known that each of a number of background factors—inheritance, early experiences, specific skills, motivational and personality influences, for example—makes a contribution to the likelihood of a particular individual becoming sufficiently eminent, creative, and original to be acknowledged a genius, but it is unlikely that it will ever be possible to use knowledge of these background circumstances to produce accurate predictions about

the likelihood of that individual's becoming a genius. Too many of the factors that lead to that happening are determined by chance or by circumstances that are beyond any individual person's control.

CONCLUSIONS

The concept of genius is hard to define or measure because genius is not simply a description of someone's qualities but a statement about the value accorded by society to the person's intellectual achievements. Individual geniuses vary enormously in their backgrounds, interests, and personalities, but most geniuses, as well as being highly intelligent, are exceptionally industrious and determined individuals who have a strong sense of direction and are willing to commit large portions of their lives to the single-minded pursuit of their goals.

(See also: GIFTEDNESS; TERMAN'S GIFTEDNESS STUDY.)

BIBLIOGRAPHY

- BERRY, C. (1990). On the origins of exceptional intellectual and cultural achievement. In M. J. A. Howe (Ed.), *Encouraging the development of exceptional abilities and talents*. Leicester: British Psychological Society.
- CSIKSZENTMIHALYI, M. (1988). Society, culture, and person: A systems view of creativity. In R. J. Sternberg (Ed.), *The nature of creativity: Contemporary psychological perspectives*. New York: Cambridge University Press.
- HAYES, J. R. (1981). *The complete problem solver*. Philadelphia: Franklin Institute Press.
- HERSHMAN, J. D., & LIEB, J. (1988). *The key to genius*. New York: Prometheus Books.
- HOWE, M. J. A. (1990). *The origins of exceptional abilities*. Oxford: Basil Blackwell.
- MURRAY, P. (ED.). (1989). *Genius: The history of an idea*. Oxford: Basil Blackwell.
- OCHSE, R. (1990). *Before the gates of excellence*. Cambridge: Cambridge University Press.
- PERKINS, D. N. (1981). *The mind's best work*. Cambridge, MA: Harvard University Press.
- SIMON, H. A., & CHASE, W. G. (1973). Skill in chess. *American Scientist*, 61, 394–403.
- SIMONTON, D. K. (1988). *Scientific genius: A psychology of science*. Cambridge: Cambridge University Press.
- TERMAN, L. M. (1925–1959). *Genetic studies of genius* (Vols. 1–5). Stanford, CA: Stanford University Press.

MICHAEL J. A. HOWE

GESELL, ARNOLD (1880–1961)

Dr. Arnold (Lucius) Gesell was born on June 21, 1880 in Alma, Wisconsin, son of Gerhardt Gesell and Christine Giesen Gesell. He died in New Haven on May 29, 1961. He has been described (Pesamanick, 1960, p.241) as "one of the great pioneers of child development."

Dr. Gesell graduated from the Stevens Point Normal School in Wisconsin in 1899. After two years at the University of Wisconsin he transferred to Clark University where he became a student of G. Stanley Hall. He received his Ph.D. from Clark in 1906.

Following an invitation from Lewis M. Terman, he moved to California as a professor of psychology and pedology at the State Normal School of Los Angeles. At this time, according to his own statement in his autobiography (Gesell, 1951, p.128), "I was overtaken by a strange subdued kind of restlessness, a vague sense of unpreparedness for a task that was taking shape in my mind. I wished in some way to make a thoroughgoing study of the developmental stages of childhood. But with all my training I lacked a realistic familiarity with the physical bases and the psychological processes of life and growth. To make good this defect I would have to study medicine. Five years later (1915) I received my medical degree from the School of Medicine at Yale University."

In 1911 Dr. Gesell was appointed assistant professor of education in Yale's newly formed education department. This appointment allowed him to pursue his medical studies. In 1911 he established the Yale Psycho-Clinic, which, after several name changes, became the Yale Clinic of Child Development. When he received his M.D. he was made a full professor.

Beginning with a concern for defect and deviation, he adopted the point of view that if one were to understand fully the child's intelligence and the level of his development, one must start with a study of infancy—a relatively uncharted territory at that time. Later in his career he described behavior characteristics of all ages, from birth through sixteen years.

DEFECT AND DEVIATION

Dr. Gesell's primary scientific concern throughout his professional life was the child's mind and a charting of the patterned and predictable way that the mind developed. However, one of his earliest and major contributions, although less well remembered today, was his outstanding work in helping to establish public school classes for retarded children, especially in Connecticut.

In 1915 Arnold Gesell was appointed to be the first school psychologist in the United States. In this position, his primary responsibility was to conduct mental examinations throughout the State of Connecticut, especially in rural communities, to check on what should and could be done for retarded girls and boys. Twenty-three of the sixty-four publications that he authored between 1915 and 1926 dealt with the topics of defect and deviation. (For this extensive bibliography, see Ames, 1989, pp. 299–302.) His major thesis was that it was the responsibility of the public to help in the care and education of those children with less than normal endowment. He emphasized that, although the exceptional child was not considered to be a legitimate public-school problem a generation earlier, society was gradually taking over responsibility for teaching children with special problems. Classes for retarded children, which Dr. Gesell did so much to encourage, continued effectively until the new notion of mainstreaming came along.

DEVELOPMENTAL SCHEDULES—THE GESELL/YALE NORMS

In spite of his major contribution of helping to establish special classes for retarded children in the public schools, Dr. Arnold Gesell is probably better known for his establishment of the Gesell (currently known as the Gesell/Yale) norms of infant and child development. Working against the widely held belief that children's intelligence, as well as the rest of their behavior, is almost completely malleable and determined more by the way they are treated than by their genetic inheritance, Gesell established the fact that behavior does indeed develop in a patterned, measurable way.

He measured behavior in four fields: motor, adaptive, language, and personal-social. In *Infancy and*

Human Growth (Gesell, 1928, p.142) he noted that "The term *adaptive behavior* approximates the term *intelligence* but cannot be made strictly equivalent." He established these four fields because it was always his position that mind manifests itself in virtually *everything* the child does (in contrast to Jean PIAGET, who restricted himself chiefly to the child's verbalization). Thus in 1945, in his book *How a Baby Grows* (Gesell, 1945, p.8) he noted, "Intelligence is only one phase of the baby's mind. Our first task is to understand the ways in which the whole baby grows. Mind and brain grow together."

He described the use of his tests in these four fields as *developmental diagnosis*, and the evaluation reached was identified as a Development Quotient (DQ). Dr. Gesell's customary definition of developmental diagnosis was that it was the application of graded functional tests of behavior to determine the maturity and the integrity of the central nervous system. Thus in *Infant Development* (Gesell, 1952, p.78) he noted that "Developmental diagnosis is concerned with levels and patterns of maturity which must be formulated in interpretive statements, using specific age norms for descriptive clarity."

The exact relationship between the DQ and the more familiar IQ tended to be a little ambiguous. A clear distinction between the two was made late. When school readiness was first discussed in "The Preschool Child" (Gesell, 1923, p.74) Dr. Gesell noted, "Importance of an intelligence rating: All that has been said in regard to examination and supervision of physical development applies with equal force to mental developmental. It is important to discover the individual differences and the psychological excellencies and weaknesses of the school entrant."

In the years that followed there were many general discussions of the relationship between the two measures. By 1941, in *Developmental Diagnosis* (Gesell & Amatruda, 1941, p.181) the distinction was made very clear: "Although the principle which underlies the DQ is similar to that of the IQ, there are important differences in clinical application. The DQ, unlike the IQ, is not limited to a single inclusive formula. A distinctive DQ can be derived for each of the four major fields of behavior. Any adaptation of our tests and methods which for psychometric convenience would

affix IQ's to infants is undesirable and is inadequate for the scientific study of growth processes.”

However, before that time, in both his clinical practice and his writings, the two terms were used synonymously to a large extent. This was especially clear in his work in the field of adoption. Environmentalists, like psychologists at the University of Iowa, insisted that the intelligence of any adopted child was determined by the intellectual level of the adopting family. Dr. Gesell's position was that to a large extent intellectual level was determined genetically. He also held that if infants and children were intelligent enough to be adoptable, within reason the intellectual level of the child should be matched to that of the adopting family. Thus although children were examined and evaluated in his clinical work on the basis of the Gesell norms, the *diagnosis* was often given in terms of *level of intelligence*: superior, high average, and so on.

As late as 1930, in his little-known but significant book, *Biographies of Child Development: The Mental Growth Careers of Eighty-four Infants and Children* (Gesell and others, 1939), Dr. Gesell was clearly using the two terms more or less interchangeably. The thesis of this book was that there tended to be a marked consistency of function as children grew older. That is, children who tested low average at a year or so of age tended to continue to rate low average during their first ten years of life.

In a few cases a Stanford Intelligence Test as well as a Gesell Development Examination was used, and the specific Stanford IQ (in figures) was given. However, for the most part, though the testing consisted of behavior tests in the four fields of behavior—that is, they *were* developmental tests—the results were given in terms of general intellectual level.

Thus, though Dr. Gesell indicated the difference between DQ and IQ fairly early, in actual clinical practice and in his writings the two were often used somewhat synonymously.

INTELLIGENCE

Dr. Gesell appeared to identify the child's response to the so-called adaptive tests (Cubes, Copy Forms, Incomplete Man, Formboard and so on) more closely with intelligence than he did with the other three

fields of behavior. However, at older ages, the language items correlated best with currently popular language tests of intelligence.

Identification of the DQ with the IQ is probably most effective in the first eighteen months of life, before language has made substantial progress. Thus it seems fair to say that a 40-week-old infant who pokes at a pellet, grasps a pellet with pincer prehension, gets to hands and knees and creeps (all developmentally expected behaviors) is also of at least average intelligence. A 32-week-old infant who exhibited these behaviors would be ahead developmentally and, thus, one could assume of superior intelligence as well.

As late as 1952, in one of his final publications, the book *Infant Development* (Gesell, 1952, pp. 2,3), he asks “How early can a mental diagnosis be made? Is it feasible to measure a baby's intelligence? What is the nature of the infant's mind? Are babies born alike and is it possible to estimate the influence of environment? And how can we ‘get at’ a baby's mentality? An infant cannot introspect and talk to you. He cannot attend to quiz questions such as we were routinely using with older children. If a baby could only talk how much simpler it would be to conduct a mental examination!

“Yet we could not go on the premise that all mental diagnosis must be deferred until a child reaches the age of verbal comprehension and speaks in sentences. We incline to the belief that the infant *from the beginning* has a mind which makes itself manifest in accessible signs and symptoms.”

This emphasis on mind continued throughout his lifetime. As late as 1951 in his film, “Embryology and Behavior” (Gesell et al., 1951), Dr. Gesell reiterated his customary statement: “Mind manifests itself in all aspects of behavior, that is in virtually everything the infant or child may do.”

However, as Dr. Gesell moved on toward evaluating children five and six years of age, especially in determining whether or not any given child was old enough behaviorally (or developmentally ready) to begin kindergarten, it gradually became quite evident to him that a child with a high IQ (ahead of what would be expected at his age developmentally) could at the same time be young for his age developmentally. Since the 1960s, the determination of proper placement in school became a prime consideration of Gesell work-

ers. In making that determination the distinction between the DQ and IQ has been most clearly recognized by Gesellians.

(See also: INFANCY; INFANT TESTS AS MEASURES OF EARLY COMPETENCE.)

BIBLIOGRAPHY

- AMES, L. B. (1989). *Arnold Gesell: Themes of his work* (pp. 17–29, 299–302). New York: Human Sciences Press.
- GESELL, A. (1923). *The preschool child* (p. 74). Boston: Houghton Mifflin.
- GESELL, A. (1928). *Infancy and human growth* (p. 142). New York: Macmillan.
- GESELL, A. (1945). *How a baby grows* (p. 8). New York: Harper & Brothers.
- GESELL, A. (1951). *Arnold Gesell: A history of psychology in autobiography* (p. 128). Worcester, MA.: Clark University Press.
- GESELL, A. (1952). *Infant development: The embryology of early human behavior* (p. 8). New York: Harper & Brothers.
- GESELL, A., & AMATRUIDA, C. S. (1941). *Developmental diagnosis* (p. 181). New York: Paul B. Hoeber.
- GESELL, A. ET AL. (1939). *Biographies of child development: The mental growth careers of eighty-four infants and children*. New York: Paul B. Hoeber.
- GESELL, A. ET AL. (1951). *Embryology of human behavior* (film). Chicago: International Film Bureau.
- PASAMANICK, BENJAMIN (1960). Arnold Gesell on his eightieth birthday. *Child Development*, 31 (2), 241.

LOUISE BATES AMES

GIFTEDNESS The identification of an individual as gifted typically means that the individual is considered to have extraordinary ability in some area. One speaks of gifted musicians, gifted athletes, gifted linguists. Although colloquial and general agreement exists about who is gifted in any one of these areas, considerable disagreement arises concerning the more technical definition of giftedness and how it is to be measured. This is particularly the case with respect to intelligence and the identification of who is to be considered intellectually gifted.

The definition of giftedness is intimately related to decisions about the measurement instruments used to define who is gifted. In areas other than intelligence, no standard measuring instruments exist for identifying the gifted. The initial identification of those considered to be gifted in music, art, or sports, for example, typically rests on observations and judgments of individuals who are themselves considered to have gifts in the particular area. Thus, someone gifted in music is usually identified by a person also gifted in music. This initial judgment is then subjected to the corroboration of other musicians, critics, and, ultimately, a portion of the general public. The identification of the gifted in this manner tends to occur in fields that eventually involve some form of public performance or presentation (Horowitz & O'Brien, 1985; Sternberg & Davidson, 1986).

The identification of the intellectually gifted does not usually rest on personal judgment. The intellectually gifted are most often identified by use of an intelligence quotient (IQ) test. This practice has its roots in the invention and development of so-called mental tests. Mental tests were originally designed to identify the intellectually gifted for the purpose of encouraging selective mating, in the hope that such selective mating would improve the human race. Subsequently, the mental test evolved into a standardized test that yielded an IQ. A high level of performance on the intelligence test and a high IQ became the major means for identifying the intellectually gifted individual.

Although the IQ is now the standard means by which intellectually gifted individuals are identified, its utility as a measure of intellectual giftedness is a subject of considerable debate. Because the measuring instrument used to identify the intellectually gifted individual also defines what is meant by intellectual "giftedness," the debate about the utility of the IQ also involves a debate about how best to define and identify the intellectually gifted.

The gifted person has always been the person that was seen as someone with extraordinary abilities. In early human history such individuals were likely to be viewed as possessed of special, sometimes divinely inspired abilities and capable of magical powers. These abilities and powers were evaluated relative to the rest

of the population or group of which the individual was a member. In time, as human societies developed and compartmentalized their activities, particular work roles and responsibilities were defined. Eventually, disciplines involving skills and expertise were formed. The gifted, those with the extraordinary abilities, were increasingly differentiated in terms of the skills and expertise required or valued by a community. In this sense, the definition of giftedness and the attention given to the gifted are always based on the values of a given society.

THE INTRODUCTION OF MENTAL TESTS

In the evolution of role and skill in societies, the gifted were typically equated with the "superior." By the mid-nineteenth century, prior to the invention of mental tests, a broad belief system about the mentally superior individual had developed in Western societies. The intellectually gifted person was often thought of as prone to madness or subject to periods of emotional instability. If not the mad genius or emotionally unstable person, the intellectually gifted individual was at least thought more likely to engage in odd, eccentric, or otherwise quirky behavior.

The evidence has never supported any of these ideas about the intellectually gifted individual. Indeed, evidence is suggestive of quite a different perspective on how well intellectually gifted individuals function in a wide variety of areas (Janos & Robinson, 1985). Nevertheless, the idea that someone is mentally superior or intellectually gifted may still conjure up notions of emotional and temperamental abnormality. Some modern discussions have led to the serious consideration that individuals highly gifted in the arts may be prone to experience periods of depression and mental instability to a greater degree than might be expected among those not in the arts.

In the late nineteenth century, Francis GALTON introduced the idea of making tests that would categorize people according to their individual differences. His efforts provided the basis for the development of the first mental tests and the quantitative approach (as distinct from personal judgments) to identifying the mentally superior. Galton's mental tests for individual differences were subsequently supplanted by the de-

velopment of the intelligence test and the use of an IQ.

The initial intelligence tests were devised in Paris by Théodore Simon and Alfred BINET as a diagnostic tool for identifying those children in the Paris school system who were so far below mental capability for their age as to be unlikely to benefit from regular schooling. In some ways, it is ironic that a test initially devised to screen for individuals considered to be mentally deficient should have become the prototype for the test used to identify those who are mentally normal and, in particular, those who function at a superior level and are thus designated as gifted.

Binet and Simon's test was revised by Lewis TERMAN at Stanford University and was named the Stanford-Binet. The basic psychometric characteristics of the Stanford-Binet became typical of many of the subsequently developed individually and group-administered intelligence tests for children. Items to be included in the test are chosen after exploring the pass-fail rates by groups of individuals at different ages who are considered normal. In the individually administered tests, the items are generally presented to the test taker in order of difficulty. The IQ score is derived by noting the number of items successfully completed, relative to the individual's age. The scoring of the test is set so that if an individual passes most of the items up to those appropriate for his or her age, the individual's IQ will be around 100 and thus average or normal. More failures will result in an IQ below 100, and more successes will result in an IQ above 100.

The point at which an IQ is considered high enough to classify an individual as mentally superior or gifted is based on a somewhat arbitrary decision. One can declare those individuals to be gifted who have a measured IQ that is 130 and above, 140 and above, or 150 and above. Sometimes a distinction is made between superior and gifted. Individuals with IQs between 120 and 130 or 140 may be designated as superior, with higher IQs reserved for the designation of gifted. In some instances, where a test does not involve an IQ but a direct or normalized score based on failed and passed items, the intellectually superior or gifted are identified according to a particular percentage. Thus, for example, those who fall in the top 3 percent of the distribution of the scores will be designated as gifted.

Depending on the purpose for which the identification is to be made, the percentage can be broadened (e.g., the top 5%) or narrowed (e.g., the top 1%).

In the various discussions and debates about the nature of intelligence, a variety of definitions have been suggested. One involves the notion that intelligence can be characterized in terms of its general (*g*) and specific (*s*) aspects. In this kind of scheme, one can talk about individuals who display giftedness in their general intelligence and individuals who show giftedness in a specific kind of intelligence.

ALTERNATIVES TO THE TYPICAL MENTAL TEST

The mental test, devised according to the basic prototype of the Stanford-Binet and its psychometric descendants, is the most widely used strategy for identifying intellectual giftedness, but is not without its critics. Many of the criticisms of this approach to defining the gifted are the basic criticisms of all mental tests. The standard mental tests are said to rely too heavily on specific knowledge, experience, and content. They are said to be culturally biased, so that they result in a lesser number of individuals from the culturally nondominant groups being identified as gifted.

Alternative proposals to the standard mental test are currently of two kinds. One involves the notion that mental tests provide too narrow a definition of intelligence. This alternative approach to considering intelligence and giftedness, instead of positing a single, general intelligence that can be measured by the standard IQ test or accepting the idea of *g* and *s* as describing intelligences, proposes multiple intelligences or particular domains of intelligence. For example, one theory of MULTIPLE INTELLIGENCES suggests seven: linguistic, musical, logical and mathematical, visual- and spatial-conceptual, bodily-kinesthetic, interpersonal, and intrapersonal (Gardner, 1983). Extraordinary ability or giftedness in one domain would not necessarily imply giftedness in another domain, and no *g* is posited. Identifying the gifted with respect to each of the domains will require the development of measurement instruments for each of the intelligence areas. This has yet to be done.

The second alternative to the typical mental test responds to the criticism that mental tests rely too heavily on measuring specific knowledge and experience rather than the processes involved in operation of intelligence. It is claimed that what distinguishes the gifted from the nongifted is that the gifted individual employs more efficient or different processes in handling information or dealing with problems (Sternberg, 1986). This approach requires that the gifted individual not only perform at a superior level but also approach problems differently. For example, it has been said that the gifted individual, in whatever domain, is distinguished from the nongifted individual by being particularly good at defining or finding problems to be solved (Getzels & Csikszentmihalyi, 1976).

Another approach to seeing giftedness in terms of functional processes also involves a different way of defining intelligence. A "triarchic theory" of intelligence has been proposed (Sternberg, 1985) that defines intelligence as being made up of a basic set of three components: higher-order or executive processes, which function as metacomponents; performance components, which relate to executing a task; and knowledge-acquisition components, which are the processes involved in learning new things. Giftedness, or extraordinary ability, may be demonstrated in one or more of the three components (Sternberg, 1986). Tests that are useful in detecting gifted functioning in each of these components have yet to be developed.

Another approach to defining giftedness that does not rely solely on measured intelligence involves specifying the confluence of abilities and characteristics required for the performance of gifted behavior. For example, J. S. Renzulli (1978) considered that giftedness is best defined as above-average ability (though not necessarily exceptional ability), along with some evidence of creativity and some evidence of task commitment. Further, he believes that cognitive processes have been overemphasized at the expense of knowledge and skills that are as important to giftedness.

Whether it is useful to distinguish the concept of being gifted from the concept of being creative has not always been clear. Although those who are considered gifted often engage in intellectual activities that result in novel products or solutions, research on creativity has tended to focus more on artistic performance

than on intellectual performance. The research on giftedness has included studies of both intellectual performance and artistic performance. Theoretical developments in the areas of creativity and in giftedness hold promise of drawing the thinking about these two topics closer together.

THE DEVELOPMENT OF THE GIFTED

Whether giftedness is inherited or acquired has been extensively debated. Because definitions of gifted behavior are generally understood to depend on social values and social contexts, giftedness, unlike eye color and skin pigmentation, cannot be discussed in simple terms as being inherited and as having a direct relationship to genes. Additionally, recent advances in biological genetic research point to the importance of the environment in how genes are expressed in development and in the role that genes play in human behavior. Conversely, constitutional characteristics probably contribute to an interaction with environmental context in determining the expression of gifted behavior (Gottlieb, 1992; Horowitz & O'Brien, 1985).

Early Expression of Giftedness. A major longitudinal study of intellectually gifted children was begun in the 1920s by Lewis Terman. He identified 1,000 children with IQs of 140 or more and studied their physical, emotional, and personality characteristics. This study provided the first objective demonstration that intellectually gifted children are also more likely to excel physically, to have good emotional adjustment, and to function well personally and socially. Follow-up studies have revealed that although these intellectually gifted children became productive, accomplished, and well-adjusted adults, their achievements at midlife were not necessarily extraordinary (Terman & Oden, 1959). Gender in social context appears to be an important factor in subsequent achievements of those identified as gifted early in life. (Tomlinson-Keasey & Little, 1990). Studies of young gifted individuals have also revealed that in almost every case these individuals experienced intensive environmental nurturance and encouragement of their gifts (Feldman, 1986).

Giftedness Across the Life Span. What determines whether or not the early expression of giftedness continues through childhood, adolescence, and adulthood is not clear. Evidence about the continuity

of gifted behavior across life periods implicates high levels of personal motivation, persistence, and the ability to be totally absorbed in an area of activity as factors that contribute to giftedness being an enduring characteristic of an individual's behavior and accomplishments.

The circumstances or conditions that are responsible for the appearance of giftedness beyond the period of childhood are little understood. Such individuals are sometime called "late bloomers," especially when their previous levels of achievement have been quite unremarkable. The assumption appears to be that the individual's giftedness was present in earlier periods but was just unexpressed, although no evidence exists for such an assumption. Giftedness may not be an inherent characteristic but one that is the result of a confluence of factors that may occur during different life periods. Studies of the life course of women who were identified earlier in life as gifted have demonstrated that continuity in giftedness is much affected by social context and life-course experiences (Tomlinson-Keasey & Blurton, 1992).

Studies of major contributions in different fields have shown that mathematicians and physicists typically make their most creative and important contributions in early adulthood, but major contributions by writers and artists are not necessarily confined to age periods of their lives. Highly talented musical performance rarely appears for the first time during middle or old age, but in art and writing, major talents and gifts have become evident for the first time relatively late in life. Literary expression appears to be one area in which giftedness seems capable of showing itself at many points across the life span.

THE SOCIAL CONTEXT OF GIFTEDNESS

The definition of giftedness and the identification and development of giftedness depend heavily on cultural values and the societal context. Cultural values and societal context will determine which domains will be recognized for the possibility of giftedness. For example, in societies without a strong oral tradition, oral giftedness involving the telling of complex stories or the expressive recounting of historical events is not likely to be identified, developed, or encouraged. Particular societal constraints may also limit the expres-

sion of giftedness. Individuals with the ability to be gifted in dance, for instance, are unlikely to be identified in cultures that prohibit dancing, much less to be provided with an environmental context for the development of their talent.

Social definitions of giftedness and the social valuing of giftedness influence how gifted individuals are encouraged and educated. These factors also determine whether provision is made for the gifted to receive formal education in the area of their gifts. Educational policy with respect to the gifted has been formed largely in terms of intellectual giftedness as defined by performance on an IQ test. Little consensus exists about how best to educate the gifted who are so identified. Some educators advocate keeping the gifted child in the regular classroom and enriching his or her experiences, but others advocate special classes for gifted children that permit the children to accelerate their educational progress.

In some instances, children who are identified in a particular domain will be provided with special instructional opportunities, even though the IQs of these children might not qualify them for inclusion in a general gifted educational program. Examples can often be found in instances involving children who show special mathematical talent. Children who demonstrate giftedness in the bodily-kinesthetic domain may be tapped for particular encouragement in sports. Schools rarely provide special programs for children gifted in music or art, except in the modern development of "magnet" schools that are designed to encourage individual talents in large public school systems.

CONCLUSIONS

The definition and identification of giftedness are determined by social and cultural context. In the United States and in much of Western society, the standard and most commonly accepted definition of giftedness involves performance on mental tests and is expressed as an IQ. A number of critiques of the IQ as the sole definition of giftedness have been offered. Alternative suggestions have focused on defining giftedness in terms of performance in different domains of functioning and thinking about gifted intelligence in terms of excelling in the processes involved in intelligent behavior.

The origins of giftedness are not well understood. The early expression of giftedness is typically accompanied by strong environmental nurturance. Individuals showing early signs of giftedness may or may not continue as gifted into later life. Although in some areas the gifted are more likely to make their major contributions in early adulthood, in other areas gifted achievements are not confined to a particular period of life. The major criterion currently used for selecting individuals as gifted for educational purposes is a high level of performance on an IQ test, although one does not find much consensus with respect to how most effectively to educate and nurture those who are considered gifted by this criterion.

(See also: GENIUS; TERMAN'S GIFTEDNESS STUDY.)

BIBLIOGRAPHY

- BENBOW, C. P., & STANLEY, J. C. (1983). *Academic precocity: Aspects of its development*. Baltimore: Johns Hopkins University Press.
- FELDMAN, D. (1980). *Beyond cognitive universals*. New York: Ablex.
- FELDMAN, D. (1986). *Nature's gambit: Child prodigies and the development of human potential*. New York: Basic Books.
- FOX, L. H., & WASHINGTON, J. (1985). Programs for the gifted and talented: Past, present and future. In F. D. Horowitz & M. O'Brien (Eds.), *The gifted and talented: Developmental perspectives* (pp. 197-221). Washington, DC: American Psychological Association.
- GARDNER, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- GETZELS, J. W., & CSIKSZENTMIHALYI, M. (1976). *The creative vision: A longitudinal study of problem finding in art*. New York: Wiley.
- GOTTLIEB, G. (1992). *Individual development and evolution: The genesis of novel behavior*. New York: Oxford University Press.
- GRINDER, R. E. (1985). The gifted in our midst: By their divine deeds, neuroses, and mental test scores we have known them. In F. D. Horowitz & M. O'Brien (Eds.), *The gifted and talented: Developmental perspectives* (pp. 5-35). Washington, DC: American Psychological Association.
- HOROWITZ, F. D., & O'BRIEN, M. (1985). *The gifted and talented: Developmental perspectives*. Washington, DC: American Psychological Association.

- JANOS, P. M., & ROBINSON, N. M. (1985). Psychosocial development in intellectually gifted children. In F. D. Horowitz & M. O'Brien (Eds.), *The gifted and talented: Developmental perspectives* (pp. 149–195). Washington, DC: American Psychological Association.
- RENZULLI, J. S. (1977). *The enrichment triad model: A guide for developing defensible programs for the gifted and talented*. Mansfield Center, CT: Creative Learning Press.
- RENZULLI, J. S. (1978). What makes giftedness? Reexamining a definition. *Phi Delta Kappan*, 60, 180–184.
- STERNBERG, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. New York: Cambridge University Press.
- STERNBERG, R. J. (1986). A triarchic theory of intellectual giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness* (pp. 223–243). New York: Cambridge University Press.
- STERNBERG, R. J., & DAVIDSON, J. E. (EDS.). (1986). *Conceptions of giftedness*. New York: Cambridge University Press.
- TERMAN, L. M., & ODEN, M. H. (1959). *Genetic studies of genius: Vol. 5. The gifted group at mid-life: Thirty-five years' follow-up of the superior group*. Stanford, CA: Stanford University Press.
- TOMLINSON-KEASEY, C., & BLURTON, E. U. (1992). Gifted women's lives: Aspirations, achievements, and personal adjustment. In J. Carlson (Ed.), *Cognition and educational practice: An international perspective* (pp. 151–176). Greenwich, CT: JAI Press.
- TOMLINSON-KEASEY, C., & LITTLE, T. D. (1990). Predicting educational attainment, occupational achievement, intellectual skills and personal adjustment among gifted men and women. *Journal of Educational Psychology*, 82, 442–455.

FRANCES DEGEN HOROWITZ

GIFTED AND TALENTED STUDENTS, EVALUATION OF EDUCATIONAL PROGRAMS FOR The health of a program for gifted and talented students requires routine evaluation to detect educational problems, document program achievements and shortcomings, and confirm that the programmatic approach is sound. Evaluators can help program administrators, funding agencies, and program participants determine whether a program is operating as planned, achieving its goals, and serving students as desired. By detecting and thus helping to prevent program deterioration, evaluation promotes

the “well-being” of gifted and talented education programs.

EVALUATION MODEL

A typical model of program assessment involves

1. determining goals and objectives;
2. describing the processes required to accomplish goals and objectives;
3. determining the immediate, direct, short-term effects of the program (comparing evidence of outcomes); and
4. determining the long-term effects of the program.

Determining Goals and Objectives. A conscientious effort should be made to determine the goals of the specific gifted program. Evaluators look not only at whether the program is fulfilling its goals but at whether program goals are specific, realistic, and measurable. For example, an artistic program may establish a goal of having every student perform a recital by the end of the year. An academic math and science program might select as goals having a certain percentage of the student body enter and win the Westinghouse Science Talent Search and Putnam Mathematical competitions. Goals can range from having students perform at two grade levels beyond their chronological peers to having them volunteer 400 hours to public service projects or conduct research, literary, or dramatic projects. Whatever the goal, it is important for the evaluators to recognize the program's explicit target outcomes before undertaking an evaluation of that program.

Describing the Processes Required to Accomplish Goals and Objectives. The second major step is to describe what the program is doing to accomplish its stated objectives. This requires continual monitoring of program operations, as well as discussions with administrators, teachers, and students. Observations of this nature will help shape questions about accountability and help explain why the program is or is not working. In addition, monitoring the program on a routine basis helps administrators, teachers, and students keep it on course—aimed at accomplishing the stated objectives of the program. Ideally, evaluation data can be used to improve program op-

erations and eliminate ineffective and wasteful facets of the program.

Determining the Immediate, Direct, Short-Term Effects of the Program (Comparing Evidence of Outcomes). The third step is to determine the program's immediate or direct impact. Are gifted and talented students performing at two grade levels beyond their chronological peers? Are students entering science and mathematical competitions in the numbers and at the level desired or expected? Are teachers providing a qualitatively different educational experience for gifted and talented students, in terms of depth and pace of study? Useful sources of information for this stage of the evaluation include interviews with administrators, teachers, students, and parents; observations of administrative and classroom behavior; and archival data, such as past evaluation reports, standardized test results, student and program portfolios, and local newspaper articles.

This is the stage of an evaluation that requires particular attention to validity and reliability. Construct and predictive validity are two of the most critical features of a measure. Construct validity is concerned with whether the measure accurately represents the concept measured. Predictive validity is concerned with the power of the measure to predict future behavior or outcomes. These concepts are described in depth in standard evaluation and measurement textbooks. In general, the evaluator must be sure to select an accurate and valid measure to determine whether a goal has been achieved. Such a measure can be scores on achievement tests, community service records, or a dramatic production. The measure selected should also be reliable or stable—in other words, it should provide the same information about a specific situation or set of circumstances over time (assuming nothing has changed about the situation).

Determining the Long-Term Effects of the Program. A final step is to consider the program's long-term or ultimate effect. Has the gifted and talented program contributed to the academic standing of the school in the community, the state, and the nation? Are more gifted and talented students entering and completing undergraduate and graduate degree programs? Are more gifted and talented program graduates making productive business, medical, or scien-

tific contributions? Typically, few programs maintain comprehensive or systematic longitudinal data on their graduates. Such data, however, provide one of the best and least expensive sources of useful impact information.

CREATING AN EVALUATION TEAM

The team approach is the most common and often the most effective method of conducting evaluations. An evaluation team is typically composed of individuals outside the program, including other gifted and talented program directors, practitioner-scholars in the field, teachers, and community members. Parents and students may also be recruited to join the evaluation team.

The team develops an evaluation plan, based on input from sponsors and key stakeholders in the school and community. Evaluation tasks—including interviews, observations, and reviews of records—are divided and distributed to appropriate team members, who execute them. The team assembles at various points in the process to share notes, report on progress, and suggest areas requiring additional investigation. Team members explore new leads and cross-check one another's data. The team chair is typically charged with the responsibility of producing an evaluation report. The team approach makes it possible to conduct a full-scale evaluation within the tight time and budget constraints typical of an educational program. It also ensures a balanced perspective, as team members cross-validate one another's judgments and observations, providing a continual check against individual observer bias.

GENERAL EVALUATION GUIDELINES

1. Make sure the evaluation supplies the practical information needed by the targeted audiences. Take time to identify who the key stakeholders are, including administrators, parents, teachers, and students. Find out what they want to know, and then develop a plan to address the most salient concerns.

2. Make sure the evaluation is realistic (politically and pragmatically) and cost effective. There is no point in conducting an evaluation if there is no political support for a program or if there is good reason to believe

that the findings will be hidden or altered. Care must also be taken to ensure that an evaluation effort does not significantly drain program resources.

3. Make sure that the evaluation is conducted in an ethical manner. The rights of program participants must be protected. Care should be taken to ensure privacy, freedom of information, and confidentiality (if promised).

4. Make sure the evaluation is as accurate as possible. Take the time to seek out various sources of information and cross-check data sources. Double check all figures and interview notes. Judgments about the data should be logical and reviewed by independent sources whenever possible.

For a more detailed discussion of evaluation standards and guidelines see *The Program Evaluation Standards* by the Joint Committee on Standards for Educational Evaluation (1992), Rossi's *Standards for Evaluation Practice* (1982), and the American Evaluation Association's *Guiding Principles for Evaluators*.

SPECIFIC GUIDELINES

1. Make sure program documentation exists. Program documentation should describe the program's philosophy; curriculum; staffing; financial, library, and computer resources; identification and screening procedures; and selection criteria. In addition, classroom schedules and maps of the physical layout will facilitate any evaluation.

2. Make sure you review as many relevant data sources as possible. Interviews and observations are critical. In addition, archival documentation, such as newsletters, financial reports, student letters, parent letters, past evaluation reports, newspaper articles, and many other documents provide pertinent data about the program's impact and role in the community.

3. Make sure you compare the program's stated goals with its actual performance. Does the program operate in accordance with its own philosophy (academically and in terms of governance)? Does the curriculum reflect the philosophy and goals of the school? Do the staff members appear to understand and implement the stated program philosophy? How do teachers translate the program's philosophy into practice in their teaching?

4. Make sure you describe and assess the climate. Are students engaged? Are teachers stimulating, thoughtful, and knowledgeable? Is communication between staff and administration constructive and cooperative or antagonistic and fragmented? Similarly, what is the nature and tone of communication with and among students?

5. Make sure you talk to students. The purpose of education programs for the gifted and talented is to serve students. Time should be devoted to informally interview students about their own progress (including a review of their portfolios, records, or projects) and their evaluation of the program. Data on students' academic achievement and behavior code are critical.

6. Make sure program finances are reviewed. Is the program budget sufficient? If not, why not? Is the money being used as intended? If not, why not? Is financial program planning adequate and appropriate to meet the needs of the program in the foreseeable future?

7. Make sure community and school board components are included in the evaluation. Do community and school board members support the program? How is this evidenced? Do parents participate in the program? What are the obstacles to community and board support, if any?

EVALUATION CONCEPTS

Evaluation draws on many useful concepts from anthropology to guide an examination of a program for gifted and talented students, including maintaining a nonjudgmental orientation, soliciting views from the insider's perspective, and triangulating data.

A nonjudgmental orientation requires the evaluator to suspend value judgments about a given behavior or practice. It allows the educational evaluator to describe an observed behavior or situation in detail before completing an analysis of classroom behavior in context. A judgmental orientation may lead to premature—and often inappropriate—assessments of a specific approach or behavior before there are enough data with which to interpret it meaningfully or place it in context.

A sound evaluation is grounded in the participants' views of what they are doing, what they are trying to accomplish, and what they think they have ac-

completed. When interviews and observations are grounded in the insiders' views of how a program works, findings are considered more salient and recommendations more realistic and practical. In addition, the evaluation can identify multiple program goals and contrast them with actual performance.

Triangulation is used to analyze data. It involves testing one source of information against another to strip away alternative explanations. Triangulation may help confirm or disprove information and can force the evaluator to probe further to reveal another level of information. Triangulation can help evaluators learn about the program's effectiveness and about how well it matches its stated philosophy in practice (Fetterman, 1989, 1993).

EVALUATION TECHNIQUES

Among the many techniques evaluators use, the most important are observation, interviewing, and participation.

Observation. There is no adequate substitute for direct observation in evaluation. Evaluators can observe students systematically—for example, observing whether different cultural groups interact; whether shy students become more outgoing; when a lesson plan is engaging; or whether an administrator participates in teaching or social activities associated with the program. Over time, patterns of behavior will become evident. These patterns, documented repeatedly over time, are a form of reliability. Observations can be used to establish baseline descriptions about the program, teacher and student performance, administrative support, and many other program-related features. Similarly, observations can be recorded to document change (and possibly growth) over time. Observations can be filed in a student portfolio to document student performance throughout the year, noting, for example, increasing sophistication in research projects, mastery of artistic expression, or complexity of mathematical problem solving. Observation is a natural tool; it is enhanced when guided by the concepts described above.

Interviews. Often the best way to check on or triangulate observation is simply to ask the observed participants what they were doing. A student singing

in chorus may seem to be enjoying herself. An interview with that student may reveal, however, that she hates chorus and is participating in it only because her parents insisted. Evaluation team members should interview key stakeholders and other team members to construct a useful list of questions.

Interviews should be open ended—that is, they should allow the person being interviewed to answer in as full and complete a manner as he or she deems appropriate. Closed yes-or-no questions are quick and efficient, but they are usually biased by the person conducting the interview and tend to shape the response. Closed questions are useful after initial interviews and observations suggest what the relevant questions are—starting from the insider's perspective.

Participation. In any educational evaluation, the most credible interviews and observations are conducted by evaluators who spend time with program participants. In this sense, evaluators who make several visits to a program—sharing in the daily lives of students, teachers, administrators, and parents—have a much more comprehensive picture of how a program for the gifted and talented operates than does the evaluator who simply comes in once in a year, asks a few questions, observes a few classes, and then writes a report about the program. There is no substitute for the depth of understanding gained from routine participation in classroom activity, extracurricular activity, home activity, faculty meetings, and school board meetings. Direct participation exposes the evaluator to the multiple levels and goals of program participants and affiliates. It also enhances the validity of the evaluation findings.

QUALITATIVE AND QUANTITATIVE APPROACHES

Evaluations should combine qualitative and quantitative approaches. Qualitative approaches are used to describe a program, its operation, and its effects. They are also required to establish the appropriate baseline for or context within which to interpret student performance. An accurate measure of student performance requires preentry information about student, program, and community characteristics, including cultural, social, political, and economic factors. For example, it is important to document whether the eval-

uation focuses on a program that is just starting up or on a mature program that has been operating for years. In addition, an accurate measure of how far a student has traveled requires qualitative and quantitative baseline data about the student's previous attendance patterns, grades, reading level, achievement record, and so on. Such data can also indicate whether a student is underachieving. In addition, qualitative approaches (such as ethnography, an established qualitative approach in evaluation) describe program implementation in sufficient detail to explicate the multiple goals of a program; delineate the process in which students, teachers, administrators, and parents interact (in school and at home); highlight adaptive and maladaptive social arrangements; and document program effects. Programs for gifted and talented students are particularly amenable to the use of qualitative approaches, in that creative approaches to teaching can be described and documented with a minimum of obtrusiveness.

Quantitative measures document outcomes, including student achievement over time. There are many forms of quantitative data—besides standardized tests—that document the effectiveness of programs for gifted and talented students, including the number of and frequency of participation in intellectually competitive activities—such as science fairs, Westinghouse, Putnam (math), and other engaging enterprises—as well as more mundane figures, such as attendance, books read, portraits or recitals completed, number entering postsecondary institutions, and so on. Contract systems provide useful quantifiable outcome data that are easily verifiable, concerning the extent to which students are achieving stated goals and objectives. Teacher and student survey questionnaires and rating scales can be developed that focus on specific topics.

Together, qualitative and quantitative approaches effectively document and analyze evaluation data. (See Coleman, 1985; Fetterman, 1988a; Gallagher, 1985; Howley, Howley, & Pendarvis, 1986; Kitano & Kirby, 1986; Reis & Renzulli, 1991; Renzulli, 1975; and Renzulli & Ward, 1969, for a detailed discussion about evaluating gifted and talented education programs. In addition, see Fetterman, 1988b, for a discussion about the variety of qualitative approaches to evaluation in education.)

COMMUNICATION AND PROGRAM PRACTICE

Evaluators must communicate their findings and recommendations if they are going to be effective. Verbal communications and memoranda issued throughout an evaluation keep evaluators and program personnel on track and provide a quality control on the data while the evaluation is still in progress. (By the time the final report is written, it is often too late to correct erroneous ideas that have been used to build the conceptual foundation of the evaluation findings.) Interim reports are valuable forms of communication for program participants, sponsors, and evaluators. They create benchmarks in the study and provide interested parties with findings to date, current thinking, and a progress report on the evolving direction of the evaluation. All of these interim forms of communication minimize surprises, gross errors, and related problems. The final report puts a cap on the entire experience for the evaluator, participants, and sponsor and typically includes recommendations that, if used properly, will serve as a guide to improve program practice.

INTERNAL EVALUATION

In conjunction with external evaluations, educational administrators, teachers, parents, and students should periodically conduct their own evaluations of their program. A typical internal evaluation team is composed of teachers, counselors, administrators, parents, and a student. Local experts, board members, and program officials are recruited to serve whenever possible. Team members should be encouraged to conduct informal self-appraisals on a daily or at least weekly basis, questioning and comparing what students are doing in relation to stated program goals and objectives. Systems should be developed to give regular feedback to students, teachers, administrators, and parents, including parent-teacher conferences, faculty meetings, and student performance conferences.

The concepts and techniques of internal evaluation, or self-evaluation, derive from traditional external evaluation. Internal evaluations also follow the same general model that external evaluators use. Program participants clearly can—and in many cases do—apply evaluation concepts and techniques in their daily

lives in these programs. Observation, measurement, documentation, and evaluation are all normal activities of teachers, students, parents, and administrators. Combining these concepts and techniques helps to make these activities more systematic and improves the validity and reliability of the self-evaluation.

Self-examinations and external evaluations, in addition to sharing concepts and techniques, can complement each other and help to cross-validate data from each approach. Self-evaluations help maintain an educational program's health on a daily basis; expert external evaluation is essential to an in-depth and objective understanding.

Fetterman (1989) created the concept of empowerment evaluation. This involves using evaluation to help others help themselves. One facet of this approach involves encouraging program participants to conduct their own evaluations and to share useful models and concepts with them to help them help themselves.

CONCLUSION

Evaluation is needed to learn how a program for gifted and talented students works, how effective such programs are, and how to raise their standards of quality. Formative evaluations provide a continual flow of information to program officials throughout a review to improve program practice. Scriven refers to formative evaluation as evaluation aimed at the improvement of an ongoing enterprise (see Davis, Scriven, & Thomas, 1981). Summative evaluations can complement formative evaluations. According to Davis, Scriven, and Thomas (1981), summative evaluation is conducted for an external client and its primary purpose is to report on the quality of the program for purposes other than improvement. The knowledge gained at the end of a summative evaluation culminates in a final report with a focus on policymaking. Evaluations, whether qualitative or quantitative, formative or summative, can improve program practice and student performance. They are also required to establish the utility of such approaches in education for the gifted and talented as acceleration, enrichment, and special group settings. Overall, evaluation plays an essential part in the development, maintenance, and under-

standing of educational programs for the gifted and talented.

(See also: GIFTEDNESS.)

BIBLIOGRAPHY

- COLEMAN, L. J. (1985). *Schooling the gifted*. Menlo Park, CA: Addison-Wesley.
- DAVIS, B. G., SCRIVEN, M., & THOMAS, S. (1981). *The evaluation of composition instruction*. Inverness, CA: Edgepress.
- FETTERMAN, D. M. (1988a). *Excellence and equality: A qualitatively different perspective on gifted and talented education*. Albany: State University of New York Press.
- FETTERMAN, D. M. (1988b). *Qualitative approaches to evaluation in education: The silent scientific revolution*. New York: Praeger.
- FETTERMAN, D. M. (1989). *Ethnography: Step by step*. Newbury Park, CA: Sage.
- FETTERMAN, D. M. (1993). *Evaluate yourself*. National Research Center on the Gifted and Talented. Storrs: University of Connecticut.
- GALLAGHER, J. J. (1985). *Teaching the gifted child*. Boston: Allyn & Bacon.
- HOWLEY, A., HOWLEY, C. B., & PENDARVIS, E. D. (1986). *Teaching gifted children*. Boston: Little, Brown.
- KITANO, M. K., & KIRBY, D. F. (1986). *Gifted education: A comprehensive view*. Boston: Little, Brown.
- REIS, S. M., & RENZULLI, J. S. (1991). The assessment of creative products in programs for gifted and talented students. *Gifted Child Quarterly*, 35 (3), 128–134.
- RENZULLI, J. S. (1975). *A guidebook for evaluating programs for the gifted and talented*. National/State Leadership Training Institute on the Gifted and Talented. Ventura, CA: Office of the Ventura County Superintendent of Schools.
- RENZULLI, J. S., & WARD, V. S. (1969). *Diagnostic and evaluative scales for differential education for the gifted*. Storrs: University of Connecticut.

DAVID M. FETTERMAN

GLASER, ROBERT (1921–) Robert Glaser was born on January 18, 1921, in Providence, Rhode Island. In 1942, he received a B.S. in chemistry

from the City College of New York. He earned an M.A. in experimental psychology in 1947 and, two years later, earned a Ph.D. in psychological measurement and learning theory, both from Indiana University. In addition, Glaser has received honorary doctorates from the University of Leuven, Belgium; the University of Göteborg, Sweden; and Indiana University.

Robert Glaser began his teaching career as an instructor at Indiana University, before becoming an assistant professor of psychology at the University of Kentucky, a research assistant professor at the University of Illinois, and then a senior research scientist with the American Institutes for Research in Pittsburgh. Moving to the University of Pittsburgh as an associate professor of psychology in 1956, he quickly became professor of both psychology and education. He founded the Learning Research and Development Center at the University of Pittsburgh in 1963, an institution he led over the course of an influential career.

Glaser served on a number of national and international boards and commissions and was active as a consultant to foundations and government agencies. He served as president of the American Educational Research Association (AERA), the American Psychological Association's Division of Measurement and Testing and Division of Educational Psychology, and the National Academy of Education (NAE), and as of 1993 was the chairman of the NAE Panel on the Evaluation of the National Assessment of Educational Progress Trial State Assessment Project. His numerous awards include a Guggenheim Fellowship, the AERA Award for Distinguished Research, and the American Psychological Association's E. L. Thorndike Award and Distinguished Scientific Award for the Applications of Psychology.

Robert Glaser's scholarship was extensive, and his influence on both the science of learning and the development of a field of application for this science was immense. He was the author or editor of 16 books, published over 220 articles, and served on the editorial boards of several scientific journals. A review of Glaser's publications reveals an extraordinary combination of continuity and of responsiveness to new trends and promising ideas. He began his publishing life as a student of measurement theory and human

performance. A behavioral psychologist, Glaser entered the arena of education as an advocate and analyst of programmed instruction, teaching machines, and the theory that gave birth to them. He became a leading proponent of cognitive science, espousing a view of the human being as thinker that many consider a radical departure from the behaviorist view. This major shift in scientific perspective was accomplished without an abrupt break either in scholarly productivity or in fundamental research and social commitments. Across the behavioral-cognitive line, Glaser maintained a continuing set of core questions and preoccupations. These include the nature of aptitudes and individual differences, the interaction of knowledge and skill in expertise, the roles of testing and technology in education, and training adapted to individual differences.

On each of these topics, Glaser wrote influential papers and edited widely read collections. At each stage of his career (e.g., Glaser & Bassok, 1989), he demonstrated extraordinary ability to locate the key issue, identify the seminal idea, consider the best work being done, and formulate an emerging set of problems that would mobilize and organize the work of many others. He did this in the late 1950s for programmed instruction (Glaser & Lumsdaine, 1960), showing how the basic idea introduced by Skinner could penetrate and energize all forms of education and training. He extended the principles of behavioral analysis to individualized education, first developing the notion of individually prescribed instruction as a means of organizing an entire elementary school to permit individual progress in accordance with the best learning principles of the time. Subsequently, he supported younger colleagues in developing variants and extensions of individualized education. Finally, he gave all of these programs theoretical advocacy under the label *adaptive education* (Glaser, 1977), a term that for Glaser encompassed a vision of an educational system devoted to teaching all students rather than selecting an able few for success.

During these years, Glaser took a leading position on two issues that continue to be at the forefront of educational development—testing and technology—always relating these to his central vision of educational possibilities. He introduced criterion-referenced

testing—to determine exactly what students know and can do, not how they compare with other students (Glaser, 1963). Such testing is a central requirement for any educational system that aims to adapt instruction to all students. Testing aimed at supporting instruction rather than certifying (or denying) ability remained one of Glaser's continuing interests, later informed by cognitive analysis of aptitudes and abilities. Glaser maintained an active personal and collaborative research program on instructional testing (e.g., Glaser, 1986; Glaser, Lesgold, & Lajoie, 1987).

As early as the 1960s, well before educational technology became a popular field for research and development, Glaser saw possibilities for using computers to enhance and enrich the instructional process (Cooley & Glaser, 1969). He pursued this vision in the context of newly enlarged computing capabilities, developing and studying computerized discovery laboratories that help students appreciate the structure of a field of knowledge by actively exploring it instead of trying to absorb the prepackaged messages of others (e.g., Shute, Glaser, & Raghavan, 1989).

Much of Glaser's work on cognitive analyses of learning derived from his long-standing interest in individual differences (Glaser, 1977). Early on, he saw the possibilities for understanding such differences through empirical studies of the cognitive processes involved in various kinds of aptitude tests (e.g., Gitomer et al., 1987; Pellegrino & Glaser, 1979). As those studies proceeded, Glaser became increasingly impressed with the role of knowledge differences in aptitudes (Glaser, 1984, 1991). This led to a companion line of research on the nature of expertise in specific academic disciplines and technical domains (Chi, Feltovich, & Glaser, 1981; Glaser, 1990) and on differences in how people acquire knowledge in particular domains of expertise (Chi et al., 1989; Glaser et al., 1992). Not surprisingly for someone whose theoretical and applied lines of work were always intertwined, and who was always more interested in how to improve human competence than in how to select people of already developed ability, Glaser's research on aptitudes and knowledge acquisition led him to explore the nature of learning and thinking skills and how these might be taught (Chipman, Segal, & Glaser, 1985; Segal, Chipman, & Glaser, 1985).

BIBLIOGRAPHY

- CHI, M. T. H., BASSOK, M., LEWIS, M. W., REIMANN, P., & GLASER, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, 13, 145–182.
- CHI, M. T. H., FELTOVICH, P., & GLASER, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5, 121–152.
- CHIPMAN, S. J., SEGAL, J. W., & GLASER, R. (EDS.). (1985). *Thinking and learning skills: Research and open questions* (Vol. 2). Hillsdale, NJ: Erlbaum.
- COOLEY, W. W., & GLASER, R. (1969). The computer and individualized instruction. *Science*, 166, 574–582.
- GITOMER, D. H., CURTIS, M. E., GLASER, R., & LENSKY, D. B. (1987). Processing differences as a function of item difficulty in verbal analogy performance. *Journal of Educational Psychology*, 79, 212–219.
- GLASER, R. (1963). Instructional technology and the measurement of learning outcomes: Some questions. *American Psychologist*, 18, 519–521.
- GLASER, R. (1977). *Adaptive education: Individual diversity and learning*. New York: Holt, Rinehart & Winston.
- GLASER, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 39, 93–104.
- GLASER, R. (1986). The integration of instruction and testing. In E. Freeman (Ed.), *The redesign of testing in the 21st century: Proceedings of the 1985 ETS Invitational Conference* (pp. 45–58). Princeton, NJ: Educational Testing Service.
- GLASER, R. (1990). Expertise and assessment. In M. C. Wittrock (Ed.), *Cognition and testing* (pp. 17–30). Englewood Cliffs, NJ: Prentice-Hall.
- GLASER, R. (1991). Intelligence as an expression of acquired knowledge. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 47–56). Hillsdale, NJ: Erlbaum.
- GLASER, R., & BASSOK, M. (1989). Learning theory and the study of instruction. *Annual Review of Psychology*, 40, 631–666.
- GLASER, R., LESGOLD, A. M., & LAJOIE, S. (1987). Toward a cognitive theory for the measurement of achievement. In R. R. Ronning, J. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing and measurement* (pp. 41–85). Hillsdale, NJ: Erlbaum.
- GLASER, R., & LUMSDAINE, A. A. (EDS.). (1960). *Teaching ma-*

- chines and programmed learning: A source book*. Washington, DC: National Education Association.
- GLASER, R., SCHAUBLE, L., RAGHAVAN, K., & ZEITZ, C. (1992). Scientific reasoning across different domains. In E. De Corte, M. Linn, H. Mandl, & L. Verschaffel (Eds.), *Computer-based learning environments and problem solving, Vol. 84* (pp. 345–371). Berlin: Springer-Verlag.
- PELLEGRINO, J. W., & GLASER, R. (1979). Cognitive correlates and components in the analysis of individual differences. *Intelligence, 3*, 187–214.
- SEGAL, J. W., CHIPMAN, S. F., & GLASER, R. (EDS.). (1985). *Thinking and learning skills: Relating instruction to research* (Vol. 1). Hillsdale, NJ: Erlbaum.
- SHUTE, V., GLASER, R., & RAGHAVAN, K. (1989). Inference and discovery in an exploratory laboratory. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences* (pp. 279–326). San Francisco: Freeman.

LAUREN RESNICK

GODDARD, HENRY H. (1866–1957) Henry Herbert Goddard was born in Vassalborough, Maine, in 1866. He received his master's degree from Harvard University in 1889. Goddard then moved to California, where he taught for a year at the University of Southern California. Gradually moving back East, he spent approximately six years as a high school principal in Ohio and later in Maine. He then returned to school and, encouraged and influenced by G. Stanley Hall, received a doctorate from Clark University in 1899. He next took a job at the Pennsylvania State Teachers College where he worked until 1906. He made a one-year trip to study in Germany, where he was impressed by the newly republished reports of the genetic experiments of Gregor Mendel. These became the impetus for most of Goddard's later work and research in intelligence (Hothersall, 1984; Scheerenberger, 1983).

In 1906, Vineland Training School in New Jersey opened the first U.S. research program in mental retardation, and Goddard was appointed its first director. The series of scales that Alfred BINET had developed with his student Théophile Simon had appeared in France the previous year. Goddard translated these scales and produced an English adaptation in

1908. (When Binet revised his scales in 1911, Goddard immediately translated the revisions [Goddard, 1910a, 1911a].) Intelligence testing was not of very great importance in the United States until Goddard produced his Binet-Simon test adaptation, but when this test was presented, it led to a surge of activity in the use of tests for the measurement of individual differences and general intelligence.

Goddard began a study of “feeble-mindedness” at the Vineland Training School. He pioneered a study of the school's clients using his Binet-Simon translations to do a kind of intelligent quotient (IQ) testing. A concept akin to IQ was introduced by Goddard. Binet himself was opposed to the single-score-type estimation of intelligence (Wolf, 1973). Goddard tested over 2,000 public schoolchildren in New Jersey in addition to those at the Vineland Training School. His goal was to try to develop a “reliable means of distinguishing between normal and feeble-minded” children and to “develop a reliable way of distinguishing between different levels of mental ability for all children.” In this connection, he introduced the word “moron.” Goddard's 1911 translation of the Binet-Simon test became the standard test for measuring intelligence in the United States until Lewis TERMAN's 1916 revision, which was to become *the* Stanford-Binet.

Goddard's major goal was to find “feeble-minded persons,” and he embarked on a major testing program throughout the New Jersey school systems. Among the children at Vineland, he found a girl whose family history fit his “Mendelian” model. He named this family “the Kallikaks” (Goddard, 1912a). His study of them, though lacking in appropriate research design or procedure, became the classic support for a heredity-based public policy (Goddard, 1942; Hothersall, 1984; Rosen, Clark, & Kivitz, 1976). He became thoroughly convinced that feeble-mindedness followed the rules of direct inheritance, and he urged that the “feeble-minded” (today described as persons with mental retardation) be institutionalized. He was in favor of sterilization on a case-by-case basis, but he was extremely cautious about its generalization into “sterilization laws” (Goddard, 1920; Scheerenberger, 1983). He came to feel that the teaching of the “3 Rs” was almost entirely inappropriate for children who were judged to be feeble-minded. He was in favor of “special” education programs and was adamant about using the

services of only the best-trained teachers for these students. This would mean establishing separate and special programs, and he worked for the development of such programs within the institutions designated for the feeble-minded.

In 1882 Congress had passed a law forbidding entry to the United States of “lunatics and idiots.” In 1913 Goddard was invited to Ellis Island to investigate the screening procedure. He said that “using the psychological method . . . the percentage of immigrants that would be picked out as defective would be much greater than now” (Goddard, 1917). These findings were based on primitive testing methods, lack of attention to language differences, inappropriate use of what were then very crude instruments, and so on, but Goddard felt that he had “scientific justification for restrictive quotas and immigration laws.” He was highly active in the development of tests for the military in World War I. He also joined forces with the eugenics movement based on his feeling that society should be organized in terms of intelligence.

In 1918 Goddard left Vineland and moved to Ohio to head its State Bureau of Juvenile Research. He studied intelligence on a much broader basis, and he again concluded that the major intellectual differences should be considered on both biological and racial bases. He began to question government by democracy, feeling that it could not function properly with the existing intelligence base. He proposed a system of “meritocracy,” feeling that equality was a myth and again favoring some types of eugenic practices. In 1922 he was appointed professor of abnormal and clinical psychology at Ohio State University. He remained there until his retirement in 1938.

Goddard also studied gifted children (Goddard, 1922), advocating special classes for them. He also began studying other aberrations of intelligence and became very much interested in individuals with “multiple personalities” (Goddard, 1927). During this period, he became more and more interested in the influence of environment and culture and began to feel that environmental influences (although they might be caused by the mixture of people with good intelligence and poor intelligence) might have the greatest influence in the shaping of personality (Goddard, 1946).

Goddard was the major pioneer in the development and advocacy of intelligence testing in the United

States. His translation, use, and research of the original Binet unquestionably initiated a new era in psychology, an era in which the analysis of individual differences would be based on psychometric instruments, and measured intelligence would become an essential part of educational and vocational placement. His theory that intelligence was a fixed, genetically determined characteristic in humans gave rise to many unfortunate consequences and had the effect of slowing the lines of research into the source and function of intellectual activity in humans. Goddard realized this later in life, and his opinions became more flexible.

BIBLIOGRAPHY

- GODDARD, H. (1910a). Four hundred feeble-minded children classified by the Binet method. *Journal of Psycho-Asthenics*, 15.
- GODDARD, H. (1910b). Heredity of feeble-mindedness. *American Breeders Magazine*, 4, 165–178.
- GODDARD, H. (1911a). *The Binet-Simon measuring scale for intelligence* (rev. ed.). Vineland, NJ: Training School.
- GODDARD, H. (1911b). Two thousand normal children measured by the Binet measuring scale of intelligence. *Pedagogical Seminary*, 18, 232–259.
- GODDARD, H. (1912a). *The Kallikak family: A study in the heredity of feeble-mindedness*. New York: Macmillan.
- GODDARD, H. (1912b). Feeble-Mindedness and immigration. *The Training School Bulletin*, 9, 91–94.
- GODDARD, H. (1913). The Binet tests in relation to immigration. *Journal of Psycho-Asthenics*, 18, 105–107.
- GODDARD, H. (1914a). *Feeble-Mindedness: Its causes and consequences*. New York: Macmillan.
- GODDARD, H. (1914b). *School training of defective children*. Yonkers-on-Hudson, NY: World Book.
- GODDARD, H. (1915). *The Criminal Imbecile*. New York: Macmillan.
- GODDARD, H. (1917). Mental tests and the immigrant. *Journal of Delinquency*, 2, 243–277.
- GODDARD, H. (1919). *Psychology of the normal and subnormal*. New York: Dodd, Mead.
- GODDARD, H. (1920). *Human efficiency and levels of intelligence*. Princeton, NJ: Princeton University Press.
- GODDARD, H. (1927). *Two souls in one body*. New York: Dodd, Mead.
- GODDARD, H. (1928a). *School training for gifted children*. Yonkers, NY: World Book.

- GODDARD, H. (1928b). Feeble-Mindedness: A question of definition. *Journal of Psycho-Asthenics*, 33, 219–227.
- GODDARD, H. (1942). In defense of the Kallikak study. *Science*, 95, 574–576.
- GODDARD, H. (1946). What is intelligence? *Journal of Social Psychology*, 24, 51–69.
- GODDARD, H., & HILL, H. F. (1911). Feeble-Mindedness and criminality. *Training School Bulletin*, 8, 6.
- HOTHERSALL, D. (1984). *History of psychology*. New York: Random House.
- ROSEN, M., CLARK, G. R., & KIVITZ, M. S. (EDS.). (1976). *The history of mental retardation: Collected Papers, Vol. I., Sect. 5*. Baltimore, MD: University Park Press.
- SAMELSON, F. (1982). H. H. Goddard and the immigrants. *American Psychologist*, 37, 1291–1292.
- SCHEERENBERGER, R. C. (1983). *History of mental retardation*. Baltimore, MD: Paul H. Brookes.
- WOLF, T. H. (1973). *Alfred Binet*. Chicago: University of Chicago Press.

HENRY LELAND

GOLDSTEIN, KURT (1878–1965) Kurt Goldstein, eminent neurologist and neuropsychologist, was born on November 6, 1878 in Kattowitz (Kato-wice), Upper Silesia, then a part of the German Empire. He died on September 19, 1965, in New York City in his 87th year. When he was a young schoolboy, the family moved to Breslau where he received a classical German secondary school education. After a year of study of philosophy and literature at the University of Heidelberg, Goldstein returned to Breslau to study medicine and received the M.D. degree in 1903. Both his own inclinations and the influence of Carl Wernicke, then professor of psychiatry in Breslau, led him to a lasting preoccupation with aphasic disorders and psychiatric disturbances.

Following some years of postgraduate study in neuropathology and clinical neurology, Goldstein joined the staff of the neuropsychiatric department of the University of Königsberg. In 1914, on the eve of World War I, he moved to the University of Frankfurt. In 1916, he organized and became the director of a hospital for the treatment and rehabilitation of brain-injured military personnel, a position that he held through the 1920s.

Goldstein was appointed professor of neurology and director of the Neurological Institute of the University of Frankfurt in 1919. Another move in 1930 took him to Berlin as director of the neurological department of a large general hospital with a professorial appointment in the university. After the rise to power of the National Socialists in Germany, he was forced to flee to Holland in 1934. Supported for a year by the Rockefeller Foundation, he wrote his magnum opus, *Der Aufbau der Organismus* (English translation: *The Organism*, 1939). The volume is a definitive statement of his views on the methodological approaches and conceptualizations required for the fruitful study of human brain function and behavior. In 1935, at the age of 57, he emigrated to the United States. Over the following thirty years, he engaged in clinical practice, held professorial appointments, and was active as a teacher and researcher in New York and Boston.

SCIENTIFIC AND SCHOLARLY CONTRIBUTIONS

Throughout the early years of his career, from 1906 to 1916, Goldstein was a prolific contributor to conventional neurology, writing on topics in neuropathology, APHASIA, comparative neurology, and localization of function in the brain. During and directly after World War I, his thinking about the significance of the symptoms of nervous disease and the behavioral consequences of brain lesions underwent a radical change. Among the factors contributing to this change in approach and conceptualization were his experiences in the evaluation and care of brain-injured soldiers, his association with psychologists of the Gestalt school, and his own holistic philosophic predilections. Goldstein rejected the mechanistic approach to the nervous system and its derangements that, by comparing specific symptoms and behavioral deficits with the anatomic findings at postmortem examination, had sought to elucidate the functions of different regions of the brain and their connections. He could scarcely quarrel with those clinical applications of the approach that had had notable success in relating symptoms to focal brain disease and that had established clinical neurology as an “exact” medical specialty. He forcefully disagreed with its basic assumptions of a connectionist reflex physiology and of discrete areas of the brain

with specific functional properties that, when diseased, manifested themselves in specific symptoms.

Instead, Goldstein proposed a holistic theory of brain function, a central tenet of which was that even a limited lesion produces a change in total brain activity. The behavioral counterpart of this total change in brain function was an impairment in what he called the “abstract attitude.” This extremely broad concept covered not only the capacity for abstract reasoning (as its name suggested) but also a rather bewildering variety of other behavioral characteristics including initiative, foresight, self-awareness, behavioral flexibility, and the ability to analyze a complex situation into its components. Goldstein interpreted these diverse behavioral deficits as expressions of a single basic capacity or Grundstörung. Largely rejected by neurologists, the concept was adopted for a time by clinical psychologists in their efforts to understand the deviant behavior of patients with brain disease, particularly those with frontal lobe lesions. In due course, however, this single-principle approach was abandoned as it became evident that the concept of the “abstract attitude” was too comprehensive and multifaceted to serve as a useful explanatory principle.

A more lasting contribution was Goldstein’s interpretation of certain symptoms as being defensive reactions on the part of patients—to protect themselves from confusion, feelings of failure, and loss of self-esteem. Thus, under certain circumstances, patients who show apathy, excessive rigidity, lack of concern, or facetiousness may be exhibiting these behaviors to cope with disabilities and to avoid the painful awareness of their cognitive incompetence. Conventional neurology generally viewed these deficits as simply manifestations of focal brain disease; for example, lack of concern and facetiousness were part of the “frontal lobe syndrome.” In doing so, however, it had to ignore the instability of these deficits and the oscillations between apathy and hyperexcitability or between facetiousness and obvious anxiety shown by these patients. Goldstein’s views were more in accord with clinical reality.

Another of Goldstein’s important contributions involved methodology. He insisted that noting a performance simply as being intact or defective or scoring a test response as right or wrong were not adequate descriptions of a patient’s behavior. Qualitative analysis

of a patient’s performances was required if one were to gain some understanding of the nature of the cognitive processes underlying the performances—whether or not they were marked “right” or “wrong.” Thus, a successful performance might still reflect the operation of disturbed cognitive processes that managed, somehow, to achieve the required overt response.

AN ASSESSMENT

Goldstein’s sharp attack on established approaches and concepts, coupled with his personal egocentricity, made him something of an outsider in mainstream neurology. In addition, he seemed inconsistent from time to time in his views on some issues, for example, aphasic disorders. Yet his ideas foreshadowed and influenced the subsequent development of certain trends of thought in both clinical neurology and neuropsychology.

As has been mentioned, Goldstein’s concept of the abstract attitude was not widely accepted. Nevertheless, it did point out that focal brain lesions are likely to affect overall cerebral functioning and thus alter a wide spectrum of behavior. The current emphasis on the functional significance of neural networks and the decline of discrete areal localization is more in accord with Goldstein’s approach than with the hypothesis of “centers” that was dominant in his day. Additionally, the emphasis on “executive functions” in contemporary cognitive neuropsychology incorporates many of his specifications of what constitutes the abstract attitude. His interpretation of certain symptoms as being, at least under some circumstances, protective reactions—rather than direct behavioral expressions of a focally diseased neural substrate—is now deeply ingrained in neuropsychological thinking and practice. Finally, in neuropsychological assessment, Goldstein’s insistence on the necessity of analyzing the distinctive characteristics of task performance is clearly reflected in the current emphasis on “process analysis.”

BIBLIOGRAPHY

- GESCHWIND, N. (1964). The paradoxical position of Kurt Goldstein in the history of aphasia. *Cortex*, 1, 214–224.
- GOLDSTEIN, G. (1990). Contributions of Kurt Goldstein to neuropsychology. *Clinical Neuropsychologist*, 4, 3–17.

- GOLDSTEIN, K. (1919). *Die Behandlung, Fürsorge und Begutachtung der Hirnverletzten*. Leipzig: Vogel.
- GOLDSTEIN, K. (1927). Die Lokalisation in der Grosshirnrinde. In A. Bethe (Ed.), *Handbuch der normalen und pathologischen Physiologie*, Vol. 10. Berlin: Springer.
- GOLDSTEIN, K. (1939). *The organism*. New York: American Book Company. (Original German version: *Der aufbau des organismus*. The Hague: Nijhoff, 1934.)
- GOLDSTEIN, K. (1942). *Aftereffects of brain injuries in war: Their evaluation and treatment*. New York: Grune & Stratton.
- GOLDSTEIN, K. (1948). *Language and language disturbances*. New York: Grune & Stratton.
- GOLDSTEIN, K. (with M. Scheerer). (1941). Abstract and concrete behavior: An experimental study with special tests. *Psychological Monographs*, 53 (Whole No. 239).
- RIESE, W. (Ed.) (1967). *Autobiography of Kurt Goldstein*. In E. G. Boring & G. Lindzey (Eds.), *A history of psychology in autobiography*, Vol. 5. New York: Appleton-Century-Crofts, 1967.
- SHAKOW, D. (1966). Kurt Goldstein (1878–1965). *American Journal of Psychology* 79, 150–154.
- TEUBER, H.-L. (1966). Kurt Goldstein's role in the development of neuropsychology. *Neuropsychologia*, 4, 299–310.

ARTHUR BENTON

GROUP TESTS The origins of group tests of intelligence or general ability go back to the Chinese empire over 2,000 years ago. A rigorous series of examinations covering a number of broad fields helped select administrators for the vast, complex domain controlled by the emperor's court. The intelligence tests that are widely used in the United States, however, spring from a context closer in time and geography.

In the early 1900s, the work of the French psychologist Alfred BINET on the assessment of intelligence of Paris schoolchildren was transported to the United States, where it won an influential set of adherents. Some of the most prominent educational psychologists of the period helped prepare American versions of the Binet intelligence tests.

Development of group testing of intelligence received an enormous boost when the Army Alpha Test

was administered in the United States to 1.7 million World War I recruits. The Army Alpha Test was a paper-and-pencil test designed for groups of people and for testing general intelligence. It contained general knowledge, vocabulary, analogical reasoning, and mathematical questions. The Army Alpha Test reflected a new national movement in the United States to apply the principles of the emerging field of psychology to important social problems.

The use of the examinations reflected a belief that individuals differ in general intellectual abilities. Those persons with higher abilities would tend to be better in a wide variety of activities. Tests for intelligence identified people at the lower end of the ability range who might be expected to have considerable difficulty handling intellectual tasks. Test administrators also could identify individuals with superior intelligence and assign them to tasks requiring substantial capacity. The Army Alpha test was welcomed as an efficient way to collect information on large groups in a relatively short time. The use of intelligence tests in the military setting had strong proponents, who characterized such testing as a perfect application of "scientific management" of people. They characterized the arguments against the use of intelligence tests to obtain "objective information" about intellectual competencies as backward and antiscientific behavior.

In the United States, the early use of group tests of intelligence in military, business, and educational settings quickly led to the identification of issues that continue to be important today. Among these issues are the following:

1. What are appropriate uses for intelligence tests?
2. What does the word "intelligence" really mean?
3. What is appropriate content for an intelligence test?
4. Are the tests fair for individuals from different groups?
5. Do the tests have appropriate technical qualities?

USE OF INTELLIGENCE TESTS IN SCHOOLS AND BUSINESSES

By the end of World War I, the use of the Army Alpha test was widely, although not universally, viewed as a success, and it led to the diffusion of the group

IQ test into schools and businesses. Arthur Otis, who had made major contributions to the Army Alpha test, used it as a model for the Otis Group Intelligence Scale, which he offered for both school and business use. Other paper-and-pencil tests, designed for group administration, developed within the emerging field of commercial test publishing, which drew its professional staff, in part, from the people who had served in the testing units of the World War I military services. The developers of the group tests built on the work associated with the individually administered IQ tests that were growing in popularity (see INDIVIDUAL TESTS; STANFORD-BINET INTELLIGENCE TEST).

Some of the group tests that became popular in the decade after World War I remain in use. One of the usages of group tests of intelligence is in association with tests of educational achievement. School districts have administered either group or individual tests of intelligence, such as the Stanford-Binet or the Wechsler Intelligence Scales for Children, along with achievement tests such as the Stanford Achievement Test Series, the Iowa Tests of Basic Skills, or the California Achievement Tests to students throughout their elementary and secondary school years. Because the intelligence test scores and the educational achievement scores were obtained for the same students, teachers and other educational staff could compare performance on the two kinds of tests.

This practice led to characterizing some students as “overachievers,” that is, students who scored higher on the achievement tests than other students with the same IQ scores. On the other hand, students who scored lower on their achievement tests than other students with similar IQ scores were viewed as “underachievers.” In some instances the overachievers received special praise for doing well in their subject-related work despite their somewhat low IQ scores. On other occasions, teachers worried that perhaps the “overachieving” students were pushing themselves too hard. The teachers encouraged the students to slow down and not tax themselves unduly! Researchers have not considered such interpretations of intelligence test scores appropriate since the 1960s in part because there is a high degree of similarity (i.e., high statistical correlation) between IQ tests and tests of language and mathematical achievement.

WHAT'S IN A NAME?

Even though groups tests of intelligence remain popular, the use of the term *intelligence* or *mental ability test* has declined considerably. The tests have been renamed to avoid what some psychologists call the *surplus meaning* associated with the word *intelligence*. In the 1920s and 1930s, some proponents of intelligence tests thought of them as measures of innate ability. Individuals and groups who scored low on the tests supposedly had not only less proficiency on specific skills but also less underlying ability. Although testing psychologists and other experts generally abandoned this point of view, the general public has persisted in the belief that IQ tests by whatever name measure innate abilities and resist influence from formal education and other learning experiences. By that line of reasoning, a person with low IQ test scores is unfortunate inasmuch as no emphasis on study and critical analysis has much value. Testing experts, however, do not hold this discouraging view today.

Because the unwarranted interpretation that group IQ tests measured innate ability stood in the way of reasonable inferences (predictions) from scores on the tests, the producers of such tests attempted to discourage those interpretations by changing the names of their group IQ tests. A clear example of this phenomenon is the test currently called the Otis-Lennon School Ability Test (OLSAT). Traceable directly to the World War I Army Alpha Test, OLSAT had originally been known as the Otis Group Intelligence Scale. This name was changed to the Otis-Lennon Mental Ability Test and then to the Otis-Lennon School Ability Test.

Today, the major group tests of intelligence all avoid the label *intelligence*. Thus OLSAT is the test that is normed and sold with both the Stanford Achievement Test Series and the Metropolitan Achievement Tests. The Cognitive Abilities Test plays the same role for the Iowa Tests of Basic Skills as does the Test of Cognitive Skills for the California Test of Basic Skills and the California Achievement Tests.

Although the group intelligence tests have dropped the word *intelligence*, no such change has occurred for the widely used individual intelligence tests. The leaders in the field are the various Wechsler tests. The Wechsler Intelligence Scale for Children and the

Wechsler Adult Intelligence Scale are two major products which compete with the Stanford-Binet Intelligence Scale. (See WECHSLER SCALES OF INTELLIGENCE.)

THE CONTENT OF GROUP TESTS OF INTELLIGENCE

The content of widely used tests of intelligence (also known as *tests of cognitive ability* or *school ability*) is fairly consistent across the testing batteries produced by different test publishers. The tests contain both verbal (vocabulary, reasoning, general information) and nonverbal (figural analogies and similar kinds of questions) components and yield both total scores and subscores. The questions included cover skills and content that can be acquired in a variety of ways, including independent reading and critical analysis, as opposed to specific coursework.

The following excerpt from the publisher's catalogue for the OLSAT describes the domain covered by these group tests of intelligence:

OLSAT is based on the premise that to learn new things, students must be able to perceive accurately, to recognize and recall what has been perceived, to think logically, to understand relationships, to abstract from a set of particulars, and to apply generalizations to new and different contexts. These processes are measured through performance on test items with pictorial, verbal, figural, and quantitative content.

Each of the group tests marketed to schools is offered with multiple levels; thus the nature of the questions becomes more sophisticated for the students in the upper-school years.

FAIRNESS OF GROUP TESTS OF INTELLIGENCE

The issue of the fairness and appropriateness of group tests of intelligence recurs persistently. An unfortunate but important event in the dialogue about fairness occurred in the years after the World War I testing, when data on recruits from the Army Alpha Test were used to support the argument that people from different national backgrounds differed in innate ability. Some prominent psychologists and others used

the conclusion from the 1920s about national differences in ability as a basis for trying to limit immigration from the countries in which people were supposedly not as smart. Today this point of view has no more adherents among testing experts than does the society of scientists who believe that the earth is flat, but in the period following World War I, the perspective was fairly widespread. Faced with a nonsensical conclusion—that the typical inhabitants of some nations were simply smarter than people unlucky enough to live somewhere else—many thoughtful people wisely discredited the whole line of evidence that led to these conclusions. Unfortunately they also rejected the idea that intellectual abilities could be measured with tests.

In fact, studies have shown that the degree of a person's exposure to the language and content of a test of intelligence is an important factor in determining the meaningfulness and appropriateness of the results of that test. This situation is quite clear when a group to be tested in the United States contains some members with little facility in English and only limited exposure to mainstream American life. If a person cannot understand the language used in a test, the test cannot accurately represent the individual's skills. Such a situation was the case for the individuals taking the Army Alpha Test. Analyses of the performance of American army recruits with different national ancestries showed that the groups who had arrived in the United States most recently scored lowest. This finding probably resulted from their limited English proficiency and lack of familiarity with American culture.

With respect to intelligence tests used in schools, the issue of fairness has been the subject of much controversy, especially the use of tests in the selection of students for special programs. Teachers often use intelligence tests to determine whether students require special assistance. Findings of disparate impact on particular racial and ethnic groups have led to judicial or other limitations on both group and individually administered tests of intelligence. A related controversy has developed over the use of group tests of intelligence to place students in various enrichment programs. School officials commonly rely on superior performance on tests of "school ability" for placement in a course or program for gifted and talented students.

In some instances, criticism of the use of group tests of intelligence relies exclusively on evaluations of the proportion of students from different groups who are selected. In other cases, critics seek evidence that the selections made on the basis of the tests are consistent with other information about students (e.g., grades in school) and about their readiness or need for special programs. With this latter approach, differences among groups comprise a question for further investigation, but some differences are reasonable as long as they are consistent with other information about the students and groups involved.

TECHNICAL QUALITY OF GROUP TESTS OF INTELLIGENCE

The primary issue in the evaluation of group tests of intelligence or school ability is their usefulness in attaining particular educational or social goals. To be useful, a group intelligence test must demonstrate sufficient reliability so that the possibility of obtaining valuable information exists. If a test does not produce consistent results, it is simply of no use as a measurement tool, and no other qualities that the test may appear to have are relevant.

If a test of intelligence has adequate consistency, that is to say, RELIABILITY, it is then appropriate to study its potential contribution to decisions. The evidence examined when weighing the usefulness of a test is known as *validity evidence*. A common example of validity evidence is the comparison of the results on intelligence tests with other information about the same group of people. For example, if a school plans to use an intelligence test to select students for a special enrichment program that will expose them to challenging reading and critical analysis of experiments or complex arguments, those responsible for the program can look to see whether students who score high on the test also do well in the program of study. If no such relationship exists, then use of the test to select students is highly questionable.

Whether a test of intelligence is valuable for a particular use is a determination that may need to be made several times, if several purposes are to be served or if a group consists of persons with diverse backgrounds. Suppose that a significant number of nonnative speakers of English make up part of the testing

group. Results for a general population of test takers may not apply to them, and special validity evaluation studies with groups of nonnative speakers may be in order. Information on test VALIDITY accumulates over time. Researchers should ask questions about validity, however, even before the first operational use of a test. When only one opportunity exists to be in a particular program, the students cannot average out their results. They are either chosen or not chosen. The fact that the test may be revised in the future does not help the individual who is not selected.

When evaluating tests of intelligence, professionals typically apply a set of guidelines known as the *Standards for Educational and Psychological Testing*. These rules describe what test makers must do to ensure the technical quality of tests. Test publishers typically also follow the guidelines in the *Code of Fair Testing Practices in Education*, a plain language version of the standards that focuses on preserving the rights of test takers. The code indicates to students and others what they may expect from test makers.

CHANGES IN THE FIELD OF GROUP INTELLIGENCE TESTING

The 1980s and 1990s have been a time of substantial theoretical and research work that has the potential to alter dramatically the nature of these tests. Work on multiple intelligences, for example, is directed toward broadening our conception of the abilities that we define as intelligence (see MULTIPLE INTELLIGENCES, THEORY OF). Similarly, direct efforts to understand what strategies people employ to engage in intelligent behavior are leading to educational interventions that may raise the intellectual level of all participants. The goal of this work is to recognize that improvements in abilities increase effectiveness in school, work, and other aspects of life. Just because some abilities come under the label *intelligence* is no reason not to work to improve these abilities.

(See also: QUICK MEASURES OF INTELLIGENCE.)

BIBLIOGRAPHY

AMERICAN EDUCATIONAL RESEARCH ASSOCIATION/AMERICAN PSYCHOLOGICAL ASSOCIATION/NATIONAL COUNCIL ON

- MEASUREMENT IN EDUCATION (1985). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- CARROLL, J. B. (1982). The measurement of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence* (pp. 29–120). New York: Cambridge University Press.
- JOINT COMMITTEE ON TESTING PRACTICE (1988). *Code of fair testing practices in education*. Washington, DC: Joint Committee on Testing Practices.
- PETERSON, J. (1925). *Early conceptions and tests of intelligence*. Yonkers, NY: World Book Company.
- STERNBERG, R. J., & DETTERMAN, D. K. (EDS.). (1986). *What is intelligence?* Norwood, NJ: Ablex.
- YERKES, R. M. (ED.). (1921). Psychological examining in the United States Army. In *Memoirs of the National Academy of Sciences*, Vol. XV. Washington, DC: Government Printing Office.

JOHN FREMER

GUILFORD, J. P. (1897–1988) Joy Paul Guilford was born March 7, 1897, on a farm near Marquette, Nebraska. He attended high school in Aurora and graduated as valedictorian. During his final year of high school he took “normal training,” through which he was introduced to psychology. Considered too young to teach, he spent a year at home, but taught the next two years.

Guilford entered the University of Nebraska in 1917. After a brief stint in the Signal Corps during World War I, he reentered the university in the fall of 1919. His mentor and sponsor was Winifred F. Hyde. Guilford was director of the psychology clinic for two years, during which time he developed firm convictions regarding the reality of individual differences among people. His first publication (1925, with Hyde) was based on the use of tests in the classification of chemistry students. After more teaching, including a new course in abnormal psychology for which he prepared by reading intensively, he finished his A.M. at Nebraska in 1924; through the efforts of Hyde and others, he was offered an assistantship at Cornell to study under E. B. Titchener.

At Cornell, Guilford became aware of “points of view,” for which Titchener was well known; in addi-

tion, Kurt Koffka was a visiting professor, and Harry Helson, with whom Guilford began a long-lasting friendship, was a newly arrived instructor. A major influence was a course in psychophysics under Karl Dallenbach. Guilford changed his minor from physiology to mathematics, with a second minor in education.

By the time he finished his dissertation, Guilford had published five articles. During the 1927–1928 academic year, Hyde announced her resignation from her post at Nebraska, and Guilford was invited to teach there as an associate professor. Guilford’s first major text, *Psychometric Methods*, was published in 1936. This established him as an authority in the field, especially as he had integrated psychophysics, psychometrics, and measurement, and included a chapter on FACTOR ANALYSIS.

While at Nebraska (1928–1940) he published six books and fifty-five articles; of the latter, his wife, Ruth, was his coauthor on six. His development of tests, primarily in the area of personality, grew during this period. In his *Autobiography* (1967), he notes that his wife helped him in his early research; when his first efforts were rejected by a well-known publisher, she created Sheridan Supply Company to meet the need. She oversaw the publication and distribution of most of his tests for the rest of his professional career, aided in the later years at Sheridan Psychological Services by their daughter, Joan.

After teaching for two summer sessions at the University of Southern California, he accepted an offer of a professorship in 1940. At Southern California, Guilford’s second year was interrupted by John Flanagan’s call to join him in the Army’s classification program, reviving an earlier interest he had had in the utility of aptitude measures. Beginning as a major as director of Psychological Research Unit # 3 at Santa Ana Army Air Base, he led research in selection and classification of aircrew trainees. Many existing tests were modified and many more were developed under his guidance; these were analyzed by factor analysis, in the first large-scale application of that technique (Guilford & Lacey, 1947). After several subsequent postings, Guilford was separated from military service with the rank of colonel early in 1946; he had returned to teaching the previous November. *Fundamental Statistics in Psychology and Education* had been published in 1942.

THE STRUCTURE-OF-INTELLECT (SOI) MODEL

Problems in the definition of intelligence, left unsolved at war's end, were addressed in a research program funded mostly by the Office of Naval Research, with additional support from the National Science Foundation and the U.S. Office of Education. The series of monographs, *Reports from the Psychological Laboratory*, began with a study of reasoning (Guilford et al., 1950) and continued until Number 43, issued in 1969. Creativity was the second project undertaken (Guilford et al., 1951).

Guilford's commitment to scientific strategy was emphasized by the practice of publishing one report detailing the hypotheses and description of tests for a study, followed a year later by the sequel reporting the administration of tests and analysis of results. Prepublishing the hypotheses and putative tests precluded revisions of the test battery; only if a test proved disastrously unreliable was it omitted from the announced battery, an event that turned out to be extremely rare. Also, the labor-intensive procedures for computing correlations, extracting factors, and rotating them made alternative analyses unfeasible.

The SOI model of intelligence evolved during the early 1950s. As Guilford (1956) presented it, the model specified that each ability had three components: the psychological operation, the content of the material dealt with, and the product configuration. Among the operations were memory (later divided into retention and recoding), cognition, evaluation, and divergent and convergent productive thinking. The contents initially presented were semantic, symbolic, and figural (later divided into visual and auditory), with behavioral soon added. The configurations, or forms of products, were considered to be units, classes, relations, systems, transformations, and implications. All in all, the consideration of all possible combinations of such descriptions as "cognition of semantic relations" led to the expectation of $5 \times 4 \times 6 = 120$ separate abilities. Each person, theoretically, had some of each, and some had more than others. (See also STRUCTURE-OF-INTELLECT MODEL.)

Research tended to follow generally held concepts of cognitive psychology—evaluation, judgment, planning, fluency, flexibility, and problem solving—always

building on previous findings regarding factors that should be considered in defining additional constructs. The first report based on aptitudes derived specifically from the SOI model was concerned with the distinction between the operations of cognition and convergent production in symbolic content (Guilford et al., 1961).

Subsequent factor-analytic studies followed the SOI-based strategy, investigating systematic subdivisions, for example, divergent production, semantic evaluation, and figural transformation. Measures of known constructs thought to facilitate performance on the tasks for new constructs were generally included in the battery with ostensible measures of the new (Merrifield, 1964). In the space of the extracted factors, rotations based on previous information were made to establish the locations of "marker vectors" for the known constructs. Remaining factors, with the "known" partialled out, were to be rotated to orthogonal simple structure as evidence of the hypothesized new constructs. Thus the canon of parsimony was observed.

CRITICISM OF MODEL AND METHODS

When computerized procedures became available some years later, the Guilford reports provided ready opportunity for others to experiment with new data-analytic techniques; some critics challenged the psychological inferences from the model on the basis of their findings, even though their data-analytic assumptions were not always met by the designs Guilford had formulated and the data he and his colleagues had analyzed some years earlier.

Carroll (1993) presents reanalyses purporting to show factorial solutions of greater clarity and consistency than those presented by Guilford and his students; but his technique of triadic analysis is not sufficiently sensitive to the extensive use of accumulated information regarding the various measures used. Specifically, the Guilford batteries for analysis were never designed for hierarchical analysis; in that context, the appearance of second- and third-order factors arises from construction errors in the measures, perhaps because insufficient attention was paid to setting the minimal level of competency in the facilitating skills. As in simpler statistics, one should not expect a

technique to produce useful results from data wherein its assumptions are not met.

SOI APPLICATIONS TO EDUCATION

In Japan, the SOI model had come to the attention of Takeya Fushimi of the Research Institute of Education for Brilliant Children. By 1970, Fushimi had initiated an extensive program of training for children younger than 12, based on the processes and contents of the SOI model, and incorporating a developmental component relevant to ages 2 through 12. Instructors for the various aptitudes were specially trained, and the children “stimulated” by aptitude-related exercises twice each week.

The International Society for Intelligence Education was formed in 1977 with Guilford as its first president. He continued in that capacity until his death a decade later. His student Mary Meeker played a major role in these developments; through her SOI Institute, she initiated application of assessment and training along SOI lines in the United States. In 1988, the ISIE published *An Odyssey of the SOI Model*, containing Guilford’s autobiography (1967) and other papers of interest, including an extensive bibliography compiled by Ruth B. Guilford.

BIBLIOGRAPHY

- CARROLL, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Oakleigh, Victoria, Australia: University of Cambridge.
- COMREY, A. L., MICHAEL, W. B., & FRUCHTER, B. (1988). J. P. Guilford, 1897–1987. *American Psychologist*, 43(12), 1086–1087.
- GUILFORD, J. P. (1936). *Psychometric methods*. New York: McGraw-Hill.
- GUILFORD, J. P. (1942). *Fundamental statistics in psychology and education*. New York: McGraw-Hill.
- GUILFORD, J. P. (1956). The structure of intellect. *Psychological Bulletin*, 53, 267–293.
- GUILFORD, J. P. (1967). *J. P. Guilford*. In E. G. Boring & G. Lindzey (Eds.), *A history of psychology in autobiography* (Vol. 5). New York: Appleton.
- GUILFORD, J. P. (1988). Afterthoughts on the structure of intellect. In A. Chiba (Ed.), *An odyssey of the SOI model: Autobiography of Dr. J. P. Guilford* (pp. 87–108). Tokyo: International Society for Intelligence Education.
- GUILFORD, J. P., COMREY, A. L., GREEN, R. F., & CHRISTENSEN, P. R. (1950). A factor-analytic study of reasoning abilities. I. Hypotheses and description of tests. *Reports from the Psychological Laboratory*, No. 1. Los Angeles: University of Southern California.
- GUILFORD, J. P., & HYDE, W. I. (1925). A test for classification of students in chemistry. *Journal of Applied Psychology*, 9, 196–202.
- GUILFORD, J. P., & LACEY, J. (1947). *Printed classification tests* (AAF Aviation Psychology Research Program Reports, No. 5). Washington, DC: U.S. Government Printing Office.
- GUILFORD, J. P., & MERRIFIELD, P. R. (1960). The structure-of-intellect model: Its uses and implication. *Reports from the Psychological Laboratory*, No. 24. Los Angeles: University of Southern California.
- GUILFORD, J. P., MERRIFIELD, P. R., CHRISTENSEN, P. R., & FRICK, J. W. (1961). Some new symbolic factors of cognition and convergent production. *Educational and Psychological Measurement*, 25, 659–682.
- GUILFORD, J. P., WILSON, R. C., CHRISTENSEN, P. R., & LEWIS, D. J. (1951). A factor-analytic study of creative thinking. I. Hypotheses and description of tests. *Reports from the Psychological Laboratory*, No. 4. Los Angeles: University of Southern California.
- MEEKER, M. N. (1969). *The structure of intellect: Its uses and interpretations*. Columbus, OH: Charles E. Merrill.
- MERRIFIELD, P. R. (1964). Facilitating vs. differentiating components of creativity. *Journal of Educational Measurement*, 1, 103–107.
- MICHAEL, W. M. (1988). In memoriam. *Educational and Psychological Measurement*, 48, ii.

PHILIP R. MERRIFIELD

GUTTMAN, LOUIS E. (1916–1987)

Louis E. Guttman, founder and scientific director of the Israel Institute of Applied Social Research and professor of social and psychological assessment at the Hebrew University of Jerusalem, was ranked among the sixty-two outstanding social scientists of the twentieth century (Deutsch, Platt, & Senghaas, 1971). Among his major contributions to social research are the integration of theoretical formulations and the definition of concepts with the design of the observations and analysis of the data. Guttman contributed to the whole system and to each individual aspect separately.

The third of five children, Louis Guttman was born in Brooklyn, New York, in 1916. His parents were Jewish immigrants from the Ukraine in the Russian Empire. His family moved to Minneapolis, Minnesota, where Guttman received all of his three degrees—B.A., M.A., and Ph.D.—in sociology. These degrees were awarded by the University of Minnesota in 1936, 1939, and 1942, respectively. Psychology attracted Guttman as well, and his coursework in it was equivalent to a major in this field. Graduate study made him realize the importance of mathematics in statistics, so he returned to the study of mathematics. This enabled him, while still a graduate student, to clarify and formalize intuitive—and often mathematically erroneous—techniques of data analysis that were beginning to pervade sociology and psychology. Some of this was later published in his “Outline of Statistical Prediction” (Guttman, 1941). Here lay the foundation of Guttman’s later work on scale analysis, reliability theory, factor analysis, facet theory, and the development of new methods in nonmetric data analysis.

Guttman was introduced to factor analysis during a predoctoral fellowship at the University of Chicago, where he met and came to admire L. L. THURSTONE. It is said that Guttman became upset because of Thurstone’s inadequate way of dealing with the estimation of communalities and so developed his own theorems about communalities, which are still part of factor analysis theory. Guttman’s doctoral thesis, originally proposed as a study of social status, made original contributions to matrix algebra in general and specifically to the matrix algebra of factor analysis.

In 1941, Guttman, then an associate professor in the department of sociology at Cornell University, joined the newly formed Research Branch of the U.S. War Department as an expert consultant to the secretary of war. The purpose of the Research Branch was to provide the Army command with facts about the attitudes of enlisted men on a variety of questions from job satisfaction to demobilization plans after the war. The staff of the Research Branch was made up largely of a group of dynamic young men and women, most under the age of 30 and just starting their careers in social research. Louis Guttman was one of that group. Their work resulted in the famous four volumes that have come to be known as *The American Soldier*, edited by Samuel Stouffer. One of the major contributions of

the Research Branch was Guttman’s development of scale analysis, or the *Guttman scale*. A scale may be defined as “a system of grouping or classifying in a series of steps or degrees according to a specified standard of relative size, amount, importance, perfection, etc.” A perfect Guttman scale exists when knowing any individual subject’s score on a test as a whole enables one to reproduce each of that subject’s individual answers. Guttman-type scales and their structures have been much used, particularly by sociologists, and have contributed immensely to the theory and use of multivariate tests (tests that are used to assess several independent variables and therefore contain a large number of items).

Originally, Guttman’s technique involved the *Scalogram Board*, a device that he developed at the Research Branch. The Scalogram Board was a wooden device, consisting of movable wooden strips with holes for small balls. It allowed the shifting of the rank order of both respondents and of question categories. It was developed long before the existence of modern computers and was used for many years in sociological and psychological studies.

A central theme in Guttman’s work was that measurement is not merely the assignment of numbers but the construction of structural theory. His involvement with factor analysis and other multivariate (multiple-factor) problems led him to nonmetric structural ideas, such as the simplex, circumplex, and radex (see RADEX THEORY), and to the development of facets for definitional structures. It became clear that these developments could not be useful for large-scale multivariate data sets without good nonmetric computing routines. Guttman first devised gadgets that worked in a crude fashion in a three-dimensional space, and then developed coordinate systems that were more sophisticated. Breakthroughs in computer design and programming made possible the creation and analysis of large correlation matrices, and resulted in a series of highly efficient nonmetric programs, pioneered by Jim Lingoes at the University of Michigan and resulting in the Guttman-Lingoes Nonmetric Program series (Guttman, 1968; Lingoes, 1973).

Throughout his life, Guttman emphasized the need for social theory and research to serve the public. In 1947, just a year before the establishment of the state of Israel, Guttman moved to what was then Palestine

on a Social Science Research Council fellowship. There, he applied the experience gained in World War II to establish a research unit for what became the Israel Defense Forces. Guttman headed the Psychological Research Unit, which conducted studies among soldiers and civilians of morale, food rationing, radio listening, and anticipated problems of demobilization. In 1956, this unit became the Israel Institute of Applied Social Research, which after his death was renamed the Louis Guttman Israel Institute of Applied Social Research.

The aim of the institute is to advise and counsel government offices as well as other public institutions in matters relating to research in the fields of social psychology, sociology, and related disciplines and to provide vital information on topics related to Israeli society. These include the Continuing Survey of Social Indicators; research on social policy, immigration, drug and alcohol abuse, human development and behavior genetics, mass media, economic behavior, education, values, and culture; and—of particular significance—application of Guttman's contributions to methodology and theory construction.

Guttman's contributions to the field of intelligence are evident in his posthumous seminal article entitled "Two Structural Laws for Intelligence Tests," which appeared in the journal *Intelligence* in 1991 (Guttman & Levy, 1991). In it, Guttman presents his definitional system, developed over the previous twenty-five years and stated first in Gratch, 1973, p. 37. Guttman observed—as had many psychologists before him—that correlations between any combination of pairs of mental tests almost always are found to be positive. On the basis of this set of observations, Guttman proposed a definition that specified the *range* and the *domain* for items for intelligence tests: "Any item belongs to the universe of intelligence test items if and only if its domain asks about an objective rule, and its range is ordered from very right to very wrong with respect to that rule" (Gratch, 1973, p. 37; Guttman & Levy, 1991).

This was followed by Guttman's "First Law of Intelligence Tests," which states: "If any two items are selected from the universe of intelligence test items, and if the population observed is not selected artificially, then the population regressions between those

two items will be monotone and with positive or zero sign" (Gratch, 1973; Guttman & Levy, 1991).

Although a bit too complicated for the nonspecialist to follow, the two statements above provide the formal definition of intelligence testing and the rationale for the phenomenon of positive correlations in accordance with this definition. Examples supporting the first law can be found in the literature, including Guttman's own writings (1965, 1967a, 1967b, 1970, 1991).

Guttman's "Second Law of Intelligence Tests" concerns the relative sizes of the correlation coefficients among intelligence test items and specifies a two-facet design based on the distinction between *kind of content* and *degree of complexity* of ability tests. From his insight into the interrelationships among different tests, further basic contributions by Guttman followed: the structural theories of the *simplex*, *circumplex*, and *radex* (SEE *Radex Theory*). The radex structure (developed in 1954) formed a basic element in his theoretical view of the structure of intelligence, abilities, and attitudes. The radex theory and the resulting geometric structures were made possible by *facet theory* and *facet design* and by the development of appropriate computer techniques for data analysis, particularly *Smallest Space Analysis* (SSA) (Guttman, 1968; Lingoes, 1973) of correlation matrices.

Guttman's contributions have received worldwide recognition. His awards and honors include the Rothschild Prize for Social Science (1963), the Outstanding Achievement Award from the University of Minnesota (1974), the Israel Prize in the Social Sciences (1978), the Educational Testing Service Award for Distinguished Service to Measurement (1984), and the (posthumous) Helen Dinerman Award from the World Association of Public Opinion Research (1988). Guttman was an elected member of the Israel Academy of Sciences and Humanities and the American Academy of Arts and Sciences, the first nonresident president of the Psychometric Society, and a member of a number of international organizations.

BIBLIOGRAPHY

- CANTER, D. (Ed.). (1985). *Facet theory: Approaches to social research*. New York: Springer Verlag.
- DEUTSCH, K. W., PLATT, J., & SENGHAAS, D. (1971). Condi-

- tions favoring major advances in social science. *Science*, 171, 450–459.
- GRATCH, H. (1973). *Twenty-five years of social research in Israel*. Jerusalem: Jerusalem Academic Press.
- GUTTMAN, L. (1941). The quantification of a class of attributes: A theory and method of scale construction. In P. Horst (Ed.), *The prediction of personal adjustment* (pp. 319–348). New York: Social Science Research Council.
- GUTTMAN, L. (1945). A basis for analyzing test-retest reliability. *Psychometrika*, 10, 255–282.
- GUTTMAN, L. (1946). The test-retest reliability of qualitative data. *Psychometrika*, 11, 81–95.
- GUTTMAN, L. (1947). The Cornell technique for scale and intensity analysis. *Educational and Psychological Measurement*, 4, 179–190.
- GUTTMAN, L. (1950). The basis for scalogram analysis. In Stouffer, S. et al. (Eds.), *Measurement and prediction: Studies in social psychology in World War II* (pp. 91–121). Princeton, NJ: Princeton University Press.
- GUTTMAN, L. (1953). Image theory for the structure of quantitative variates. *Psychometrika*, 18, 277–296.
- GUTTMAN, L. (1954a). A generalized simplex for factor analysis. In *Proceedings of the Fourteenth International Congress of Psychology*.
- GUTTMAN, L. (1954b). Some necessary conditions for common-factor analysis. *Psychometrika*, 19, 149–161.
- GUTTMAN, L. (1954c). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences*. New York: Free Press.
- GUTTMAN, L. (1954d). The principal components of scalable attitudes. In P. F. Lazarsfeld (Ed.), *Mathematical thinking in the social sciences* (pp. 216–257). Glencoe, IL: The Free Press.
- GUTTMAN, L. (1955a). The determinacy of factor score matrices with implications for five other basic problems of common-factor theory. *British Journal of Statistical Psychology*, 8, 65–81.
- GUTTMAN, L. (1955b). Reliability formulas for noncompleted or speeded tests. *Psychometrika*, 20, 113–124.
- GUTTMAN, L. (1957a). Empirical verification of the radex structure of mental abilities and personality traits. *Educational and Psychological Measurement*, 17, 391–407.
- GUTTMAN, L. (1957b). Introduction to facet design and analysis. *Proceedings of the Fifteenth International Congress of Psychology—Brussels* (pp. 130–132). Amsterdam: North Holland.
- GUTTMAN, L. (1958). What lies ahead for factor analysis? *Educational and Psychological Measurement*, 18, 497–515.
- GUTTMAN, L. (1960). The matrices of linear least-squares image analysis. *British Journal of Statistical Psychology*, 13, 109–118.
- GUTTMAN, L. (1965). The structure of interrelations among intelligence tests. *Proceedings of the 1964 Invitational Conference on Testing Problems*. Princeton, NJ: Educational Testing Service.
- GUTTMAN, L. (1967a). The nonmetric breakthrough for the behavioral sciences. *Proceedings of the Second National Conference on Data Processing* (pp. 495–510). Information Processing Association of Israel.
- GUTTMAN, L. (1967b). A psychological design for a theory of mental abilities. In D. N. Jackson and S. Messick (Eds.), *Problems in human assessment* (pp. 438–446). New York: McGraw-Hill.
- GUTTMAN, L. (1968). A general nonmetric technique for finding the smallest coordinate space for a configuration of points. *Psychometrika*, 33, 469–506.
- GUTTMAN, L. (1969). Integration of test design and analysis. In *Toward a theory of achievement measurement* (pp. 53–65). *Proceedings of the 1969 Invitational Conference on Testing Problems*. Princeton, NJ: Princeton University.
- GUTTMAN, L. (1970). Integration of test design and analysis. *Proceedings of the 1969 Invitational Conference on Testing Problems*. Princeton, NJ: Educational Testing Service.
- GUTTMAN, L. (1971). Measurement as a structural theory. *Psychometrika*, 36, 329–347.
- GUTTMAN, L. (1977). What is not what in statistics. *The Statistician*, 26, 81–107.
- GUTTMAN, L. (1980). Integration of test design and analysis: Status in 1979. *New Directions for Testing and Measurement*, 5, 93–98.
- GUTTMAN, L. (1982a). Facet theory, smallest space analysis, and factor analysis. *Perceptual and Motor Skills*, 54, 491–493.
- GUTTMAN, L. (1982b). What is not what in theory construction. In R. M. Hauser, D. Mechanic, & A. Haller (Eds.), *Social structure and behavior* (pp. 331–348). New York: Academic Press.
- GUTTMAN, L. (1985). The illogic of statistical inference for cumulative science. *Applied Stochastic Models and Data Analysis*, 1, 3–10.
- GUTTMAN, L. (1988). Eta, Disco, Odisco, and F. *Psychometrika*, 53, 393–405.

- GUTTMAN, L., & GUTTMAN, R. (1963). Cross-cultural stability of an intercorrelation pattern of abilities. *Human Biology*, 33, 53–60.
- GUTTMAN, L., & LEVY, S. (1991). Two structural laws for intelligence tests. *Intelligence*, 15, 79–103.
- GUTTMAN, L., & SCHLESINGER, I. M. (1967). Systematic construction of distractors for ability and achievement test items. *Educational and Psychological Measurement*, 27, 569–580.
- GUTTMAN, L., & SCHLESINGER, I. M. (1969). Smallest space analysis of intelligence and achievement tests. *Psychological Bulletin*, 71, 95–100.
- LINGOES, J. C. (1973). *The Guttman-Lingoes nonmetric program series*. Ann Arbor: Mathesis.
- SHYE, S. (Ed.). (1978). *Theory construction and data analysis in the behavioral sciences*. San Francisco: Jossey-Bass.
- STOUFFER, S. A., GUTTMAN, L., SUCHMAN, E. A., LAZARUS, P. F., STAR, S. A., & CLAUSEN, J. A. (1950). *Studies in Social Psychology in World War II: Vol. 4. Measurement and prediction*. Princeton, NJ: Princeton University Press.

RUTH GUTTMAN

H

HALSTEAD, WARD CAMPBELL (1908–1969) Ward Campbell Halstead was born on December 31, 1908, in Sciotoville, Ohio, to parents of Scotch-English descent. He attended Miami University in Miami, Ohio, from 1925 to 1927 and Ohio University, from which in 1930 he received a B.A. At Ohio State University he received an M.A. in 1931. He then attended Northwestern University, where in 1935 he received his Ph.D. in psychology. Halstead's doctoral dissertation was concerned with the effects of cerebellar lesions on postrotational nystagmus in pigeons. Although he used animals for some of his experimental studies throughout his career, Halstead's stronger interest was in the area of human rather than animal brain-behavior relationships.

To pursue studies of the behavioral effects of brain lesions in human beings, Halstead moved to the University of Chicago, where he spent his entire professional career. Halstead's first appointment at the University of Chicago was as a National Research Council Fellow (1935–1936). He went on to become an instructor in experimental psychology in the Department of Medicine (1936–1939); an assistant professor in the Departments of Psychology and Medicine and associate member of the Otho SA Sprague Memorial Institute (1939–1943); an associate professor in the Departments of Psychology and Medicine (1943–1946); and a professor in the Departments of Psychology and Medicine (1946–1969).

At the University of Chicago, Halstead established a professional relationship with two neurological surgeons, Percival Bailey and Paul Bucy. These physicians agreed to provide Halstead with descriptions of brain lesions in patients on whom they had operated and to assist in making these patients available to Halstead for psychological evaluation. With this arrangement, Halstead established in 1935 what probably was the first full-time laboratory in the world for the study of human brain-behavior relationships. Since little relevant work had been done to evaluate intellectual and cognitive functions in persons with brain lesions as compared with control subjects, Halstead began by performing naturalistic observations of these patients in their everyday living settings. Although he conceptualized intelligence as being much broader than the types of abilities routinely measured with the intelligence tests of that era, throughout his entire career he viewed the brain as the organ of intelligence. In fact, Halstead felt that intelligent behavior was an expected manifestation of normal brain functions, as expressed by appropriate adaptive abilities, in both human beings and subhuman species. This conceptualization led to the development of his theory of "biological" intelligence, which was related to but also differed in a number of respects from concepts that had developed from neurological observations, and from concepts that stemmed from clinical observations.

Halstead described these various approaches to in-

telligence in his 1947 book, *Brain and Intelligence*. Although Halstead and many other investigators had published results of evaluations of persons with cerebral lesions, his concept of biological intelligence can be considered the first formal theory based upon comparisons of persons with known and carefully described brain lesions as contrasted with normal subjects. Halstead's theory conceptualized intelligence as composed of four factors: (1) the Central Integrative Field factor, which represented differences in the residue of the accumulated background and experience of the individual; (2) the Abstraction factor, which was the basic, or core, ability on which intelligent behavior depended and which was represented in individuals by their differences in the capacities associated with abstraction, reasoning, and the ability to observe cues that permitted rational categorization; (3) a Power factor, which was the energizing source of intellectual functions and presumably related principally to differences across individuals in the neurochemical and metabolic aspects of brain functions; and (4) a Directional factor, which represented differential input to the brain from the environment through the various senses, followed by output from the brain in terms of directed and organized responses.

In a number of respects, Halstead's concept of biological intelligence was similar to theories proposed as early as the late 1800s, as well as to the mid-twentieth-century theory proposed by the Russian physician Aleksandr LURIA. Shortly after Halstead published his theory of biological intelligence, a contemporary Canadian psychologist, Donald HEBB, described his own theory of intelligence, which also was based on physiological concepts of brain functioning. Since the time of these three contributors, however, there have been few attempts to relate intelligence to theoretical concepts of brain functions. Halstead felt that his second most important contribution, beyond describing his theory of biological intelligence and the development of tests sensitive to brain damage, was his theoretical description, developed with Katz, of the biochemical nature of memory. Although this brain-based theory of memory has been cited by many researchers, it has not survived as a contribution to current concepts and approaches to memory.

Retrospectively, Halstead's major contribution stemmed from his close collaboration with neurosur-

geons, who described their patients' brain lesions in sufficient detail to serve as a basis for correlation of those lesions with behavioral measurements, as well as his development of standardized testing procedures for evaluating intellectual and cognitive functions that reflected those biological conditions of the brain. These tests of biological intelligence that Halstead developed broadened the professional community's awareness of the range of intellectual and cognitive abilities dependent on the brain and served as procedures both for measurement and for evaluation of biological intelligence in the individual human subject. In turn, with other scientists' additions to the test battery, the procedures became applicable for use in clinical settings. The Halstead-Reitan Neuropsychological Test Battery, as it is now known, is reported to be the most widely used battery of neuropsychological tests in the United States and probably in the world. These contributions have been the basis for identifying Ward Halstead as the father of the discipline now known as clinical neuropsychology.

Halstead also had a strong interest in the neuropsychological correlates of aging, and his work in this area led to a number of studies by his students. Certain of the tests developed by Halstead, especially those concerned with abstraction and reasoning, have shown that some cognitive abilities decline significantly with normal aging.

Halstead died of amyotrophic lateral sclerosis on March 25, 1969, at the age of 60.

BIBLIOGRAPHY

- HALSTEAD, W. C. (1947). *Brain and intelligence*. Chicago: University of Chicago Press.
- HALSTEAD, W. C. (1951). Biological intelligence. *Journal of Personality*, 20, 118–130.
- HALSTEAD, W. C. (1951). Brain and intelligence. In L. A. Jeffress (Ed.), *Cerebral mechanisms in behavior—The Hixson Symposium* (pp. 244–288). New York: Wiley.
- HALSTEAD, W. C., MERRYMAN, P., & KLEIN, B. (1963). The Halstead Index and differential aging. In R. Williams, C. Tibbitts, & W. Donahue (Eds.), *Processes of aging* (pp. 146–159). New York: Atherton.
- HALSTEAD, W. C., RUCKER, W. B., & McMAHON, J. P. (1967). Memory. *Annual Review of Medicine*, 18, 1–14.

HALSTEAD, W. C., & SETTLAGE, P. H. (1943). Grouping behavior of normal persons and persons with lesions of the brain. *Archives of Neurology and Psychiatry*, 49, 489–506.

RALPH M. REITAN

HEALTH AND INTELLIGENCE Contrary to a common stereotype, it is not true that intelligent people are sickly bookworms or frail nerds who can barely bench-press a pencil. In most cases, higher intelligence is associated with better health, but the relation depends on many variables. It is not intelligence itself that is causing better health. And in some circumstances, there is no association at all between intelligence and health.

The relation between intelligence and health is a function of three kinds of forces. First, there are the ways in which health affects intelligence. For example, many illnesses impair mental function. Second, there are the routes by which intelligence may affect health, such as by allowing a person to understand doctors' recommendations. Finally, there are the many underlying forces that influence both intelligence and health; generally speaking, these are factors that simultaneously impair both physical and mental functioning.

HEALTH EFFECTS ON INTELLIGENCE

Since the brain is a biological organ that communicates in various ways with the rest of the body, it is not surprising that the brain is impaired by many diseases. In some of these diseases, like encephalitis or syphilis, there is a direct infection of the brain. A significant newcomer in this area is HIV-infection, which can spread into the brain, producing subtle or more pronounced neurological problems; this may occur well before full-blown AIDS is diagnosed. In fact, sometimes neurologists are the first to diagnose a person's HIV infection or syphilis or other infection.

Other diseases affect the supply of oxygen or other vital substances to the brain, thus impairing intelligence. For example, cardiovascular disease gradually may choke off the blood supply to parts of the brain. Or, a stroke—caused by a clogged artery in the head or neck—can cut off blood supply in a matter of minutes. These are significant problems since cardiovas-

cular disease is by far the greatest cause of premature mortality in developed countries. Less frequently, a stroke is caused by a ruptured blood vessel in the brain, killing surrounding brain cells. Interestingly, if the stroke is small and limited, very particular and circumscribed effects on intelligence may emerge. For example, a patient may be unable to recognize faces, but may otherwise remain quite intelligent. Or a patient may have trouble with nouns but not with other parts of speech.

Generally speaking, the more involved the brain is in the illness, the greater is the mental impairment. In the elderly (above age 65), the vast majority of cases of mental impairment are caused by Alzheimer's disease, an illness in which brain cells degenerate and die, producing a distinctive tangle of cells that can be seen only upon autopsy. Its cause is unknown. A cancerous tumor can also cause tremendous damage as it spreads through the brain. Often, the cancer has metastasized from elsewhere in the body.

Head trauma—being injured in the head—is a significant cause of mental impairment in younger people as well as in the elderly. Sometimes the head injury also causes retrograde amnesia, so the person does not have any memory of the injury. Any sudden change in mental alertness, even in a young person, needs immediate medical attention.

Suffocation can cause brain damage in a matter of minutes. Common causes include choking, inhalation of vomit, drowning, and strangulation. Oxygen-deprivation brain damage can also result from a variety of lung diseases.

Anxiety and depression can interfere significantly with various aspects of intelligence, including attention and memory. Anxiety and depression are complicated psychological and psychobiological reaction patterns, often involving both environmental stress and neurochemical disruptions in the brain. Common symptoms of depression are feelings of hopelessness, sleep disturbances, thoughts of death, fatigue, concentration difficulties, persistent sadness, and eating disorders. Since a sizable portion of the population faces depression at some point in their lives, the effects on society are quite significant. In addition, if anxiety or depression leads to alcohol or drug abuse, further profound effects on both physical and mental functioning will result.

More generally, just about any of the many illnesses that affect the body's general metabolism can interfere with mental function. Commonly, these are disorders of the blood sugar level, or thyroid disease, or interference with oxygen supply (hypoxia), or failures to clear a bodily waste product—due to kidney or liver disease. Electrolyte imbalances (imbalanced ratios of salt ions) due to such conditions as diarrhea can also affect mental state. Often, these various diseases are slow to develop and so may cause a gradual decline in intelligence. If the brain has not been extensively damaged, intelligence may return to normal as the bodily condition is corrected.

On the other hand, there are some devastating diseases in which intelligence is fully retained. Strokes that damage only sensory or motor portions of the brain, or developmental conditions such as certain forms of cerebral palsy may leave a highly intelligent brain functioning within a paralyzed body. Similarly, a disease like ALS (amyotrophic lateral sclerosis, or Lou Gehrig's disease) involves a progressive deterioration of the nerves that control voluntary motor function, but leaves intelligence intact.

There is reason to suspect that prolonged surgery under general anesthesia can have subtle effects on brain functioning. For example, after their husbands undergo extensive coronary-bypass surgery, wives sometimes comment that they note subtle changes in mental functioning. Perhaps because there are rarely major, obvious impacts on intelligence, this puzzling phenomenon is insufficiently studied. It is not known whether such effects are due to the anesthesia techniques or to temporary life-support systems or to drug effects or to the stress of major surgery. Brain swelling may be involved.

A very important but often overlooked influence of illness on intelligence involves the effects of drugs prescribed as part of the medical treatment. A large number of widely used medications affect mental functioning to a greater or lesser degree. For example, commonly prescribed drugs to treat anxiety such as benzodiazepines like Valium and Xanax can produce confusion or interfere with memory. There are too many such drugs to list here, but almost everyone has encountered a warning to "be careful about driving or operating machinery while taking this medication."

Even over-the-counter cold remedies, consumed by millions of people, can have such effects. These effects are often intensified by alcohol consumption. Unfortunately, physicians rarely ask whether a medication is affecting a patient's mental functioning unless a major problem is obvious.

The question also arises as to whether certain drugs can improve intelligence. Certainly stimulant drugs like caffeine can improve mental alertness for short periods. But that is hardly an intelligence-enhancing effect. Theoretically speaking, drugs that enhance the efficiency with which brain cells communicate with each other could improve certain aspects of intelligence, but such effects are unproven.

EFFECTS OF INTELLIGENCE ON HEALTH

Are there things that smarter people do that protect or promote their health? Surprisingly, there is no solid evidence that intelligent people stay healthier or live longer than people of average intelligence, once we take into account social, cultural, and economic forces. For example, young men living in poor inner cities have high rates of cigarette smoking and substance abuse, and they have significantly shortened lifespans due to homicide and AIDS. But there is little evidence that intelligence has much of an effect here; it is overwhelmed by socioeconomic forces. In other words, social and economic factors account for such relations. Once we take socioeconomic status into account, intelligence has little if any additional effect on health.

Similarly, there is no clear evidence that highly intelligent people take better care of themselves than do people of average intelligence. (At the extremes of low intelligence, of course, people do not take adequate care of themselves.) Intelligent people are better able to understand a doctor's orders, but are not more likely to follow them, since many other considerations are involved. For example, intelligent people may forget to take their medication, just like people of average intelligence. Intelligent people may worry more about their illness, or doubt their doctor's wisdom. Or, they may find it just as difficult to quit smoking, increase their exercise, or change their diets (assuming they do not have a socioeconomic advantage).

Of course, people of below-average or immature intelligence may incur threats to their health if they are in situations which they cannot manage. For example, children cannot be expected to follow many doctor-prescribed regimens or ensure a proper diet for themselves. In the developed countries, however, public health and safety measures are so well mandated, and food supplies are so plentiful, that health effects due to low or immature intelligence alone are not a significant problem.

Contrary to some stereotypes, highly intelligent people are not especially likely to avoid sports or physical exercise. Generally speaking, most health-promoting activities that can be done by a well-off person of high intelligence can be (and are) done by a well-off person of average intelligence, and vice versa.

Forces Influencing Both Intelligence and Health. Although there are few direct health benefits of intelligence, especially as compared to the many intelligence-damaging effects of ill health, there are indeed many biological and environmental influences that affect both intelligence and health. These create an association between intelligence and health. It is a spurious correlation in the sense that efforts to improve intelligence will have few or no necessary effects on health.

One of the strongest yet most puzzling influences on health is socioeconomic status. It is well established that people of higher income and education are healthier and live longer. Part of the reason for this relation is that poor and less-educated people have less access to medical care, struggle with inferior nutrition and housing, and are more likely to use unhealthy substances such as cigarettes. Yet even above the poverty level, where nutrition and health care are adequate, higher socioeconomic status is still associated with better health.

Therefore, to the extent that intelligence is related to better education and income, it will also be associated with health. College-educated people generally have better health and longevity than high school drop-outs, and of course college graduates are generally of higher intelligence. But again, social and economic status seem key. For example, consider those cases—common until recently—in which college-educated affluent men take uneducated wives. The wives'

new socioeconomic status, not their intelligence, becomes the primary influence on their future health.

Biological Factors. Genetic and neonatal biological defects can impair both intelligence and health. For example, DOWN SYNDROME and Angelman syndrome, which result from chromosomal defects, generally produce mental retardation and a shortened life span. Many other genetically caused cases of low intelligence similarly have health effects. Other simultaneous effects on intelligence and health may come from birth traumas and many other factors that impede normal development.

Environmental Factors. Many environmental factors affect both intelligence and health. Any toxic substance that kills brain cells can also kill cells in other organs. One example is heavy metal. Perhaps most striking here is the case of lead, which poisons the nervous system, thereby affecting hearing, coordination, and muscle strength, as well as damaging the kidneys and other organs.

Although there is no doubt that lead poisoning can impair intelligence, there is controversy about the levels of lead that are dangerous. There is some evidence that even small increases in the level of lead in the blood are associated with decreases in children's IQ, but it is difficult to rule out other factors (such as poverty) that tend to be associated with lead exposure. The issue is an important one because although lead has been removed from gasoline and from food and beverage cans in the developed countries, many thousands of children are exposed to lead-based paint or to soil that has been contaminated with old leaded-paint dust. (Lead paint was outlawed in the United States in 1980 but still exists in most housing built before that date.) Some scientists have advocated screening the blood-lead levels in all at-risk children.

Mercury is an extremely dangerous heavy metal, known to produce severe mental and physical effects. The phrase "mad as a hatter" arose because hat makers suffered brain damage as they used mercury to make felt hats. Food contamination with mercury nowadays is quite rare, but sometimes occurs in seafood. Dental workers are at some risk, but there is no evidence that ordinary people are being poisoned by their dental amalgams. Manganese is another heavy metal believed to have serious mental and physical health effects. Poi-

sonings have been found in manganese miners, in glass and ceramic workers, and in Pacific islanders who live around manganese-contaminated volcanic soil; but here again, precise effects on intelligence have not yet been determined.

Any substances that deprive the body of oxygen or replace the oxygen in the blood can also, of course, have devastating effects. Such substances include gases like carbon monoxide and chlorine. Faulty space heaters (improperly used indoors) are a dangerous source of carbon monoxide, and possibly serious levels are found in the air of polluted cities. Children living in areas of heavy smog such as southern California have been shown to grow up with diminished lung capacity and a higher incidence of respiratory diseases. But effects of air pollution, if any, on intelligence have not been determined. Finally, it is worth noting that many insecticides are known to have neurological effects—that is partly why they are lethal to pests.

Malnutrition in childhood has been thought to impair both intellectual and physical development. Although this is obviously true in extreme cases, it is difficult to know if this is a widespread cause of problems, because malnourished children generally face many other social and environmental problems as well. The cause of mental and physical health problems might not be the malnutrition per se but rather lack of parental care, lack of play and education, absence of medical care, or a polluted environment. Fortunately, children who face malnutrition in early childhood can often be greatly helped by compensatory intervention programs at the preschool or kindergarten stage.

Aging. The risk of death and chronic disease like cardiovascular disease rises dramatically in old age. Many people assume that there is also a necessary fall in mental functioning. Is there a concurrent decline in intelligence? Longitudinal studies suggest that many people remain intellectually vigorous well into their 80s. If kept active, the aging brain can keep many of the skills it had in years past, but many people stop “exercising” their brains after they retire from work. In other words, although the best health habits cannot extend life much beyond age 85, many aged people can retain excellent mental functioning until they die.

The cells of the brain communicate through chemical substances called neurotransmitters, such as sero-

tonin and dopamine. In many people, significant numbers of brain cells die with age, and the production of neurotransmitters also declines significantly. Some specific skills such as recalling names seem to decline with age, even when general health remains good. But most old people can retain most aspects of their intelligence. Active use of the brain may build up new neural connections even while others are disappearing.

It should never be assumed that mental decline in an old person is natural and incurable. The problem may be something as simple as a vitamin deficiency. Other common causes of mental impairment in the elderly are depression, hearing loss, alcohol abuse, head injury, and malnutrition, all of which often can be successfully treated. On the other hand, many diseases associated with aging have significant effects throughout the body, including on the brain, and are irreversible.

There is some evidence that physical exercise is associated with high levels of mental functioning in the elderly. Such a finding makes sense because physical exercise is also associated with lower levels of depression, less drowsiness, better circulation of oxygen, and better nutrition. Vigorous sports activity does not seem necessary. Daily walking probably provides most of the achievable benefit.

Finally, since the aged are more likely to be taking drugs for various medical problems, the resulting mental impairments may be seen (erroneously) as a natural part of aging, rather than as a drug side-effect, even by the scientific community. Physicians should reevaluate the drug regimen of any aged person suffering sudden mental decline. (See also AGING AND INTELLIGENCE.)

CONCLUSION

Many biological and environmental factors impair both brain function and general bodily function, thus producing an association between intelligence and health. This is not, however, a necessary relation—there are many people of high intelligence in poor health, and conversely, many people with impairments of higher brain function who live robustly for many years. Furthermore, there are many paths by which disease and poor health can lead to declines in intelli-

gence. Unfortunately, the reverse is not so true, as there are few ways that intelligence seems to lead directly to better health.

(See also: NUTRITION.)

BIBLIOGRAPHY

- BUNKER, J. P., GOMBY, J. S., & KEHRER, B. H. (EDS.). (1989). *Pathways to health: The role of social factors*. Menlo Park, CA: Henry Kaiser Family Foundation.
- FRIEDMAN, H. S. (1991). *The self-healing personality: Why some people achieve health and others succumb to illness*. New York: Henry Holt.
- MILLS, S. A., & PROUGH, D. S. (1991). Neuropsychiatric complications following cardiac surgery. *Seminars in Thoracic and Cardiovascular Surgery*, 3(1), 39–46.
- NEEDLEMAN, H. L. (1989). The persistent threat of lead: A singular opportunity. *American Journal of Public Health*, 79(5), 643–645.
- OKAGAKI, L., & STERNBERG, R. J. (1991). *Directors of development: Influences on the development of children's thinking*. Hillsdale, NJ: Erlbaum.
- POON, L. W., MARTIN, P., CLAYTON, G. M., MESSNER, S., NOBLE, C. A., & JOHNSON, M. A. (1992). The influences of cognitive resources on adaptation and old age. *International Journal of Aging and Human Development*, 34(1), 31–46.
- RAMEY, C. T., & RAMEY, S. L. (1992). Effective early intervention. *Mental Retardation*, 30(6), 337–345.
- TERMAN, L. M., & ODEN, M. H. (1959). *Genetic studies of genius: The gifted group at mid-life* (Vol. 5). Stanford, CA: Stanford University Press.
- WILSON, J. D. (ED.). (1991). *Harrison's principles of internal medicine* (12th ed). New York: McGraw-Hill.

HOWARD S. FRIEDMAN

HEBB, DONALD O. (1904–1985) Donald Olding Hebb was one of the most important theorists in the first century of psychology. His thinking profoundly affected the understanding of perception, cognition, neuropsychology, and intelligence. In the domain of intelligence, in particular, he made a lasting mark, beginning the now-extensive research on environmental enrichment. His influence was even more sweeping, however. As testament, we need only look

to such disciplines as neuroscience and cognitive science, where the cell assembly and Hebbian learning are universally known. Furthermore, as of the early 1990s Hebb was still the only non-American ever to have been president of the American Psychological Association (in 1960).

Hebb was born in Chester, Nova Scotia, Canada, in 1904, the son of two physicians. His three siblings also chose to become physicians, but Hebb's career path was to be more idiosyncratic. He graduated from Dalhousie University in Halifax in 1925 with the goal of becoming a writer, but a variety of jobs diverted him. He then gained admission to the graduate program at McGill University in 1928 to pursue his writerly interest in Freud. He studied part-time, supporting himself by teaching and soon becoming principal of an elementary school in Montreal.

Between 1928 and 1934, Hebb completed a master's thesis on the spinal reflex and then undertook empirical work on the salivary reflex with an Ivan Pavlov student, Leonid Andreyev. This experience led Hebb to emphasize experimental evidence and to focus on the mind-brain relation. Had it not been for two events, he might have continued teaching and studying in Montreal. A diseased hip forced him to convalesce for a year, leaving him with a permanent limp, and his wife died in a car accident. Hebb then went to Chicago to work with Karl Lashley, whose emphasis on understanding the nervous system permanently shaped Hebb's theoretical approach. Following Lashley to Harvard, Hebb conducted his dissertation on early visual experience in rats, work that led him to argue strongly for the importance of innate neural organization. His doctorate was granted in 1936.

In 1937, having married again, Hebb returned to Montreal as a postdoctoral student with Wilder Penfield at the new Montreal Neurological Institute, where he studied the effects of brain surgery on intellectual function. His interest in intelligence emerged here. Hebb was especially excited by his finding that extensive damage to the frontal lobes did not produce intellectual loss as measured by standard intelligence tests. This was contrary to prevailing theoretical ideas (e.g., Lashley's "mass action") and increased his desire to bring brain and behavior together. He later said that this experience had set the main course of all his subsequent work.

Two years later, Hebb became an assistant professor at Queen's University in Kingston, Ontario, where he became increasingly intrigued by the role of experience in intellectual development. But in 1942 Lashley lured him away to work on the problem of personality and emotion in chimpanzees at the Yerkes primate laboratory in Florida (see Robert YERKES). Hebb's now-developing theory was greatly affected by his fascination with chimpanzee behavior and by his work on the relation between emotion and thought.

In 1947, Hebb returned to McGill as professor and chair, remaining in Montreal for the twenty-five years until his retirement. His most influential work, *The Organization of Behavior*, was published in 1949. This remarkable book proposed a general neuropsychological theory, a theory that he saw as "a form of connectionism" (p. xix). His most influential idea was introduced here: "Any frequently repeated, particular stimulation will lead to the slow development of a 'cell assembly,' a diffuse structure . . . capable of acting briefly as a closed system, delivering facilitation to other such systems. . . . A series of such events constitutes a 'phase sequence'—the thought process" (p. xix).

From the beginning, Hebb saw "intelligence" as having two different meanings. Intelligence A was "an innate potential, the capacity for development" (p. 294), whereas Intelligence B represented "an average level of performance or comprehension by the partly grown or mature person" (p. 294). Again in his words: "There are then two determinants of intellectual growth: a completely necessary innate potential (Intelligence A), and a completely necessary stimulating environment" (p. 302), with Intelligence B more directly inferred from behavior. He felt that much misunderstanding of intelligence stemmed from others not making (or not heeding) this distinction.

Hebb saw intelligence as bounded by both heredity and environment and limited by the lower of the two. He viewed standard intelligence tests as valuable for estimating Intelligence B but felt they could not accurately reflect Intelligence A because of their strong cultural loading. Nevertheless, even thirty years after the publication of *The Organization of Behavior*, he maintained that "the intelligence test is a tool of the greatest potential importance" (Hebb, 1980, p. 75). His perspective was much like Alfred BINET's, emphasizing

the diagnostic use of tests to locate children at educational risk and to determine how best to remedy their deficiencies.

Probably Hebb's greatest contribution to the study of intelligence grew out of a small pilot study. He took home some rats and raised them as pets, assisted by his two young daughters. He then compared their maze-learning performance to that of rats kept in laboratory cages, using the Hebb-Williams (1946) maze. The difference was striking: The home-raised rats were consistently the top performers. Hebb (1949, pp. 298–299) concluded that "the richer experience of the pet group made them better able to profit by new experiences at maturity." From this pioneering work emerged a new approach to the role of early experience in intelligence. In reviewing this research, M. C. Diamond (1988, p. 3) credits Hebb with the initiating hypothesis underlying the vast enrichment literature.

D. O. Hebb attempted to link mental abilities to their underlying neurological substrate. In his neuropsychological theory, intelligence required the initial laying down of cell assemblies followed by the development of facilitation among the assemblies. How successfully this was accomplished determined the level of intelligence of the organism, whether human or nonhuman. Hebb never dodged the difficult problems, such as the nature–nurture question or the issue of race differences in IQ, but his analysis was always informed by the realization that intelligence was accomplished by a biological system.

In 1972, Hebb retired, returning to his family farm in Nova Scotia, where he could enjoy sailing. He had accepted a position at Dalhousie University and remained an active departmental member there. His last major work was a monograph in 1980, a more philosophical look at the psychological issues that had dominated his thinking throughout his career. He had studied—and in turn influenced—most of the major domains within psychology, and he left a legacy of ideas that are still influential in the study of intelligence and beyond.

BIBLIOGRAPHY

- BEACH, F. A. (1987). Donald Olding Hebb (1904–1985). *American Psychologist*, 42, 186–187.
- DIAMOND, M. C. (1988). *Enriching heredity: The impact of the*

- environment on the anatomy of the brain. New York: Free Press.
- HEBB, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
- HEBB, D. O. (1960). The American revolution. *American Psychologist*, 15, 735–745.
- HEBB, D. O. (1972). *Textbook of psychology* (3rd ed.). Philadelphia: W. B. Saunders.
- HEBB, D. O. (1980). *Essay on mind*. Hillsdale, NJ: Erlbaum.
- HEBB, D. O., & WILLIAMS, K. (1946). A method of rating animal intelligence. *Journal of General Psychology*, 34, 59–65.
- MILNER, P. M. (1993). The mind and Donald O. Hebb. *Scientific American*, 268(1), 124–129.

COLIN M. MACLEOD

HEBB'S THEORY OF INTELLIGENCE

Central to D. O. HEBB's theory of intelligence (1949) is his fundamental concept of *cell assemblies*. Cell assemblies were described as collections of individual neurons, initially autonomous, that become connected to one another as a result of repeated simultaneous activation. When excited through sensory stimulation, the cell assembly produces a perception; when triggered by another cell assembly, it produces imagery, ideation, or a mediating process. When cell assemblies are simultaneously excited, they can become interconnected to produce increasingly complex systems of assemblies, which Hebb called *phase sequences*. It is through the creation of more and more cell assemblies and systems of assemblies—which are analogous to both the neural structures described by many present-day neuroscientists and the neural networks modeled by connectionists—that higher-order thought and intelligent behavior emerge.

After reviewing the literature on changes in intelligence that accompanied aging and brain lesions, Hebb identified two general classes of abilities, which R. B. CATTELL (1971) would later refer to as *crystallized and fluid intelligence* (see FLUID AND CRYSTALLIZED INTELLIGENCE, THEORY OF). The first class was said to be reflected in the proficiency with which an individual could employ his or her existing stock of information and established ways of perceiving, thinking, and solving problems. Aging and brain lesions appeared to have

little effect on performance on tests, such as the Stanford-Binet, that surveyed primarily this first class of abilities. The second class encompassed abilities to solve problems, usually nonverbal, that required the individual to go beyond his or her existing inventory of intellectual tools, or that necessitated a reorganization of them. This second class of abilities, unlike the first, appeared greatly affected by aging and lesions. Hebb surmised that the requisite neural tissue necessary for establishing new cell assemblies and new interconnections among them was greater than that required to exercise them once they were ensconced. This subsequently led Hebb to differentiate between two uses of the word *intelligence*.

Hebb contended that many of the historical disputes concerning the nature and measurement of intelligence were the outcome of inadequate conceptual distinctions. He saw in use two meanings of the word *intelligence*—*Intelligence A* and *Intelligence B*—that he regarded as being related to two levels of reality. He defined Intelligence A as a fully “innate potential” or a “capacity for development” (1949, p. 294). By this he meant the inherited structure of the brain and its neural metabolism, which he saw as setting the ceiling on the development of cell assemblies and their possible interconnectedness. He defined Intelligence B as the “average level of performance or comprehension” (1949, p. 294) attained by an individual at a given level of his or her development, revealing the actual, rather than the potential, level of cell assembly development and interconnectedness. Further, he contended that, although neither Intelligence A nor Intelligence B could be observed directly, Intelligence B could at least be estimated reliably by standard tests. Hebb (1980) equated Intelligences A and B with the genetic concepts of *genotype* and *phenotype*. *Genotypic intelligence* was said to be the hypothetical biological limits set on cell assembly development by an individual's DNA, independent of environment; *phenotypic intelligence* was the actual extent of development.

Although with respect to Intelligence A, Hebb was a hereditarian, in terms of Intelligence B, he was an archetypal interactionist. Consonant with his understanding of the genotype–phenotype distinction, Hebb regarded Intelligence B as the outcome of an interplay between genotype and the environments through which the individual passed during development, par-

ticularly in infancy. Moreover, he concluded not only that an individual's degree of Intelligence B was influenced by the environment but also that what came to be considered "intelligent," in the "B" sense, was itself defined by the environment. That is, the varieties of cell assemblies and the connections among them that will be dictated by tasks and valued by others were deemed by Hebb to be culture bound. In discussing the measurement of Intelligence B, he stated that

Intelligence in sense B . . . is quite measurable, with relatively short steps of inference by tests which in effect sample a cultural product, to determine the extent to which the person being tested has mastered the ideas, modes of thought and of solving problems, ways of perceiving, and the store of information characteristic of the society for which the test is standardized [1980, p. 74].

Because only Intelligence A was viewed as independent of time and place, and because it could not be measured, the idea of a culture-fair test was considered an oxymoron. By this, Hebb did not mean that genotypic intelligence had little influence on an individual's phenotypic intelligence. On the contrary, he argued that both Intelligence A and the environment were crucial: regardless of the particular environment, Intelligence A set the upper limit on an individual's Intelligence B.

With respect to the nature–nurture controversy, Hebb (1980) thought that there were two fallacies that needed to be avoided: the psychologist's fallacy, and the biologist's fallacy. Researching and writing during the heyday of behaviorism, he defined the psychologist's fallacy as the bias to suppose that "all behavioral problems are problems of learning, and to forget—or deny—the significance of heredity and constitution" (1980, p. 70). Like many psychologists, Hebb assumed that, given that there is a biological instantiation for all psychological phenomena, differences in genotype must necessarily be at least partly responsible for phenotypic variability. The biologist's fallacy, which he said was shared by many psychologists, "is to ask whether a given character is inherited or acquired" (1980, p. 72). Not only was this question viewed as nonsensical by Hebb, but so were the related questions of "How much depends on heredity?" and "How important is the environment?" For Hebb, attempting to apportion relative importance—either qualitatively or

quantitatively—was illogical. First, it was futile, because Intelligence A could not be measured. Second, declaring that one factor was more important than the other was like "saying that the area of [a] field depends more on its length than on its width" (1980, p. 72). As he was fond of remarking, both heredity and environment should be regarded as 100 percent vital.

Finally, unlike most other psychologists of his era, Hebb saw that heritability studies were of limited value in our quest to understand how genes interact with environments to produce particular intellectual abilities (1980, p. 77). In his view, because a heritability index is contingent upon the relative variability of the two factors, the most it could ever disclose was that some individuals have inherited better brains than have others. As it is more commonly understood today, no heritability index of intelligence can indicate the degree of malleability regarding phenotypic intelligence, which he considered theoretically and practically the more meaningful puzzle. Ultimately, Hebb's emphasis on the developmental interplay between heredity and environment represented his greatest contribution to our understanding of intelligence.

BIBLIOGRAPHY

- CATTELL, R. B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton-Mifflin.
- HEBB, D. O. (1942). The effect of early and late brain injury upon test scores, and the nature of normal adult intelligence. *Proceedings of the American Philosophical Society*, 85, 275–292.
- HEBB, D. O. (1949). *The organization of behavior: A neuropsychological theory*. New York: Wiley.
- HEBB, D. O. (1980). *Essays of mind*. Hillsdale, NJ: Erlbaum.

DOUGLAS A. BORS

HERITABILITY In the late nineteenth century, Francis GALTON launched research on the heritability of human traits with this inquiry:

I had long tried to gain some insight into the relative powers of Nature and Nurture . . . it occurred to me that the after-history of those twins who had been closely alike as children, and were afterwards parted, or had been originally unlike and afterwards reared to-

gether, would supply much of what was wanted. So I enquired in all directions for appropriate cases, and at length obtained a fair supply, on which an article in *Frazer's Magazine*, Nov. 1875, was written [Galton, 1974, p. 294].

The legitimacy of heritability estimates, especially of "intelligence," has been debated ever since from many different points of view. This article will restrict itself to the statistical aspects of heritability estimates. After setting the stage with a brief review of the needed statistical concepts, the basic definitions and assumptions of the "standard biometrical" model, which underlies most traditional heritability estimates, will be introduced, and some of the more popular estimates will be critically evaluated. Specific methods of estimation (e.g., least-squares or more elaborate maximum-likelihood methods) will not be covered. They are not germane to the overriding question of whether the postulated covariance model makes sense conceptually and, if it does, whether it actually fits the data. If the answer to one of these questions is no, then no amount of algorithmic ingenuity can remedy this defect.

After a brief review of some elementary concepts of analysis of variance and the notion of intraclass correlation, a systematic exposition will be given of the expected mean squares that the standard biometrical model implies for various twin types. They will then be used to define a number of derived indices, including reliability and various conventional heritability estimates. Finally, some limitations of inferences derived from heritability estimates will be noted.

COVARIANCE STRUCTURES AND MEAN SQUARES

The standard biometrical model is essentially a variant of a model discussed by R. A. Fisher (1918). The additive version of Fisher's variance component model involved three types of variables: an additive genetic variable, a dominance deviations variable, and a variable accounting for environmental influences. The term *additive* means that all gene-environment interactions are assumed to be zero. The standard biometrical model used in twin research adds a fourth variable to account explicitly for measurement error, which is harder to control in psychology than in bi-

ology, where the dependent variable may be number of eggs or body weight. Thus, the standard biometrical model stipulates four types of random variables, which will be denoted by the letters, a , d , e , and z :

- a : an additive genetic variable with variance $\text{var}(a)$
- d : an additional genetic variable accounting for dominance deviations from the additive effects with variance $\text{var}(d)$
- e : an environmental influence with variance $\text{var}(e)$
- z : a measurement error variable with variance $\text{var}(z)$.

Let y stand for the observed test score and subscript k indicate a specific twin within a family (i.e., write y_1 for the score of one twin and y_2 for the score of the other twin), and let μ denote the population mean of the variable y . Then the most comprehensive version of the standard biometrical model, Fisher's dominance model, can be expressed thus:

$$y_k - \mu = a_k + d_k + e_k + z_k, k = 1, 2 \quad (1.1)$$

where the subscripts refer to the two twins. On dropping the dominance deviation d_k , one obtains the simpler, more popular additive model as a special case:

$$y_k - \mu = a_k + e_k + z_k, k = 1, 2. \quad (1.2)$$

Since this model has fewer parameters than the dominance model, it requires less data. To save space, we write $\text{var}(x)$ for the variance of a variable x , and $\text{cor}(x, y)$ for the correlation between two variables, x and y . With these abbreviations, the covariance restrictions of the model can be stated thus:

$$\begin{aligned} \text{var}(a_k) &= \text{var}(a), \text{var}(d_k) = \text{var}(d), \\ \text{var}(e_k) &= \text{var}(e), \text{var}(z_k) = \text{var}(z), \\ \text{cor}(a_1, a_2) &= 1/2, \text{cor}(d_1, d_2) = 1/4, \end{aligned} \quad (1.3)$$

while all other variables are assumed to be pairwise uncorrelated (e.g., a_k with d_k , a_k with e_k , and d_k with z_k). In particular, the two environments, e_1, e_2 , of twins raised apart are assumed to be uncorrelated. Although it is, strictly speaking, the covariance restrictions (1.3) that render the models testable, they will here be simply referred to as model (1.1), and model (1.2).

The fact that models (1.1) and (1.2) include no terms for gene-environment interactions has engendered considerable controversy in the past (e.g.,

Hirsch, 1981; Wahlsten, 1990). Like most other simplifying assumptions, such as the strong independence assumptions in (1.3), this omission is primarily motivated by concerns for mathematical tractability. The empirical plausibility has to be evaluated separately in each case. This important caveat will reappear in the last section.

Although, in principle, one can write out variance component models for any type of family relationship, most studies have employed identical (monozygotic) and fraternal (dizygotic) twins, and the bulk of such research has dealt with twins raised together, though occasionally twins raised apart have also been studied. Since, altogether, four types of twins are dealt with, monozygotic and dizygotic twins raised together and raised apart, the standard biometrical model implies four different structures, as set out in Table 1.

To illustrate this concretely, consider monozygotic twins raised apart (MZAs). Since they are monozygotic (identical), they share the additive genetic component a and the dominance deviations d . Since they are raised apart, the standard model assumes they were raised in two entirely uncorrelated environments, e_1 and e_2 . In contrast, dizygotic twins raised together (DZTs) share the same environment, e , but are each characterized by a twin-specific additive genetic component, a_1 versus a_2 , which correlate .50 according to genetic theory, and a twin-specific dominance deviation variable, d_1 and d_2 , which correlate .25. The intuitive idea behind these expectations is that each child samples randomly one-half of his genes from each parent, so that the genes sampled by one sibling will randomly overlap with those of another sibling of the same parents.

The resulting structural equations, together with the stated covariance and independence assumptions,

can be used to predict the variances of the observed measure y and the degree of similarity of the scores y_1 and y_2 of two twins on the same measure, which is measured by "intra-class correlations." The basic tool for deriving these parameters is the familiar two-way "random effects" analysis variance (ANOVA) model covered in most elementary statistics courses. As a refresher, the needed terms and concepts are numerically illustrated in Table 2.

Part I of Table 2 summarizes the computations involved in a two-way ANOVA. In Section (a), a 3-by-2 array of 6 raw scores is given on the left along with the row and column sums. The overall mean is 8. On subtracting it from each of the 6 raw scores, one arrives at the 3-by-2 matrix of deviation scores given in the middle of Section (a). The 6-by-2 double-entry table to its right will be discussed later. In Section (b), the deviation scores are decomposed into three pieces: a row effect (which, since the grand mean is zero, is given simply by the row mean), a column effect (given by the column means), and the remainder, the "interaction effect." On squaring and summing these effects for each of the three submatrices, one arrives at three sums of squares: a sum of squares of the row effects, $SSR = 52$; a sum of squares for the column effects, $SSC = 24$; and a sum of squares of the interactions, $SS(R \times C) = 4$. The sum of these three sums of squares equals the sum of squares of the original deviation scores, the total sum of squares, $SST = 80$.

With each sum of squares, an integer, its "degrees of freedom" (df), is associated. If N is the number of rows ($N = 3$ in the example), then the original deviation scores around the grand mean have $2N - 1$ degrees of freedom, the matrix of row effects has $N - 1$ df and the matrix of column effects has 1 df. Finally,

TABLE 1
Variable structures postulated for different twin types under the dominance model (1.1)

Genetic Makeup	Environment	
	Reared Together ($e_1 = e_2 = e$)	Reared Apart ($e_1 \neq e_2$)
MZs ($a_1 = a_2 = a$)	$y_k = a + d + e + z_k$	$y_k = a + d + e_k + z_k$
DZs ($a_1 \neq a_2$)	$y_k = a_k + d_k + e + z_k$	$y_k = a_k + d_k + e_k + z_k$

TABLE 2
Numerical example of mean squares and intraclass correlations

I. General $N \times 2$ Analysis of Variance

(a) Raw Scores, Deviation Scores, Double-Entry Table:

	Raw Scores		Deviation Scores				Double Entry			
	<i>a</i>	<i>b</i>	<i>a + b</i>	<i>a</i>	<i>b</i>	<i>a + b</i>	<i>a - b</i>	means	<i>p</i>	<i>q</i>
	6	2	8	-2	-6	-8	4	-4	-2	-6
	12	6	18	4	-2	2	6	1	4	-2
	12	10	22	4	2	6	2	3	4	2
Sums	30	18	48	6	-6	0	12	0	-6	-2
Means	10	6		2	-2				2	4
Grand mean			8					Means	0	0
Sums of squares (<i>SSQ</i>)						104	56		80	80
Sum of products									24	

(b) General Analysis of Variance Structure of Deviation Scores:

<i>Total</i>	=	<i>Rows</i>	+	<i>Columns</i>	+	<i>Rows</i> × <i>Cols</i>	
-2 -6	=	-4 -4	+	2 -2	+	0 0	
4 -2	=	1 1	+	2 -2	+	1 -1	
4 2	=	3 3	+	2 -2	+	-1 1	
Sums of squares (<i>SSQ</i>):							
<i>SSQ</i>	80	=	52	+	24	+	4
In general	<i>SST</i>	=	<i>SSR</i>	+	<i>SSC</i>	+	<i>SS(R × C)</i>
Degrees of freedom (<i>df</i>):							
<i>df</i>	5	=	2	+	1	+	2
In general	$2N - 1$	=	$N - 1$	+	1	+	$N - 1$

II. Applications to Twin Data: Between and Within Sums of Squares

If, in applications to twin data, the rows refer to families (pairs of twins), then *SSR* becomes the between sum of squares, and $SST - SSR = SSC + SS(R \times C)$ becomes the within sum of squares:

	<i>Total</i>	=	<i>Between</i>	+	<i>Within</i>
<i>SSQ</i>	80	=	52	+	28
In general	<i>SST</i>	=	<i>SSB</i>	+	<i>SSW</i>
<i>df</i>	5	=	2	+	3
In general	$2N - 1$	=	$N - 1$	+	N
Mean squares:					
Mean squares	80/5	=	52/2	=	28/3
In general	<i>MST</i>	=	<i>MSB</i>	=	<i>MSW</i>

(c) Intraclass Correlations:

Double-entry table: $r_1' = \text{cov/var} = (24/5)/(80/5) = .3$

Sums of squares: $r_1' = (SSB - SSW)/(SSB + SSW) = (52 - 28)/(52 + 28) = .3$

Mean squares: $r_1' = (MSB - MSW)/(MSB + MSW)$

$= (52/2 - 28/3)/(52/2 + 28/3) = .472.$

NOTE: (a) The discrepancy between r_1 and r_1' diminishes quickly with increasing sample size N . The sums of squares of the columns marked $a + b$ and $a - b$ of the deviation score matrix (104 and 56) give twice the between and within sums of squares.

the matrix of interactions has $(N - 1)1 = N - 1$ df. On dividing each sum of squares by its associated degrees of freedom, one arrives at mean squares, which can be interpreted as variance estimates.

In Part II of Table 2 this general $N \times 2$ ANOVA model is interpreted in the context of twin data. Let rows represent families. Then the row sum of squares ($SSR = 52$ in the example) is interpreted as a sum of squares between families (or twin pairs), denoted SSB . The difference $SST - SSB$, which equals $SSC + SS(R \times C)$, are then the within-family variations, denoted SSW . The degrees of freedom also add, so that SSW has $2 + 1 = 3$ df in the example. On dividing SSB and SSW by their respective df, one arrives at the mean squares between ($MSB = 52/2$) and mean squares within ($MSW = 28/3$).

On inspecting the columns labeled $a + b$ and $a - b$, one finds that SSB and SSW can be obtained, more simply, by squaring and summing the row sums and row differences of the original deviation score matrix. This observation is useful for deriving the expected values of the mean squares associated with the various twin types set out in Table 3 for the general dominance model (1.1). On setting $\text{var}(d) = 0$, one obtains the mean squares under the more restrictive additive model (1.2).

The between and within mean squares add up to $2\text{var}(y)$ ($= 2[\text{var}(a) + \text{var}(d) + \text{var}(e) + \text{var}(z)]$) for all four twin types (MZT, MZA, DZT, DZA), so that, in effect, the total variance, $\text{var}(y)$, is simply redistrib-

uted in different proportions over the two types of mean squares under the standard model (1.1).

DERIVED INDICES

Reliability. Here we disregard the conceptual difficulties that Charles SPEARMAN's indeterminate factor model bequeathed on classical true-score theory, or CTT (e.g., Steiger & Schönemann, 1976; Wilson, 1928) and, following standard practice, simply adopt the conventional definition of "reliability" (r_{yy}) as the ratio of systematic over total variance:

$$r_{yy} = [\text{var}(y) - \text{var}(z)]/\text{var}(y) = 1 - \text{var}(z)/\text{var}(y) \tag{1.4}$$

To illustrate this for the MZTs, from Table 3a one finds that their systematic variance is given by $(MSB - MSW)/2$ (where MSB and MSW stand for mean squares between and within, respectively). Hence, in this case, under both models:

$$r_{yy} = (MSB - MSW)/2\text{var}(y) = (MSB - MSW)/(MSB + MSW).$$

Some authors (e.g., Plomin & Bergeman, 1991) interpret the within-twin measurement error variable z_k in minor variants of the standard model (1.1) as "nonshared environment." Since the standard biometrical model does not include any other exclusively within-twin variable, such an interpretation leads to two equally unpalatable choices: Either one must

TABLE 3
Expected mean squares under the dominance model (1.1)

	<i>Twins Raised Together</i>	
	<i>Between Pairs</i>	<i>Within Pairs</i>
MZTs	$2\text{var}(a) + 2\text{var}(d) + 2\text{var}(e) + \text{var}(z)$	$\text{var}(z)$
DZTs	$1.5 \text{var}(a) + 1.25\text{var}(d) + 2\text{var}(e) + \text{var}(z)$	$.5\text{var}(a) + .75\text{var}(d) + \text{var}(z)$
	<i>Twins Raised Apart</i>	
	<i>Between Pairs</i>	<i>Within Pairs</i>
MZAs	$2\text{var}(a) + 2\text{var}(d) + \text{var}(e) + \text{var}(z)$	$\text{var}(e) + \text{var}(z)$
DZAs	$1.5 \text{var}(a) + 1.25\text{var}(d) + \text{var}(e) + \text{var}(z)$	$.5\text{var}(a) + .75\text{var}(d) + \text{var}(e) + \text{var}(z)$

NOTE: Setting $\text{var}(d) = 0$ gives the mean squares under model (1.2).

assume that the dependent variable is perfectly reliable or one finds that the nonshared-environment variance and the measurement-error variance are inextricably confounded. Neither alternative seems especially attractive. While in theory it is of course possible to postulate two different within-twin variables, one representing measurement error and the other representing nonshared environment, and to render both variances estimable by imposing strong enough covariance conditions, this does not seem to have been done, and it may not be easy to do in practice.

Intraclass Correlations. The mean squares in Table 3 are all that is needed to define heritability estimates, and they are indispensable for stringent tests of models (1.1) and (1.2). Traditionally, heritability indices have been stated in terms of less-informative similarity indices called "intraclass correlations." For pairs, they can be interpreted simply as regular product-moment correlations computed over arrays in which each twin pair has been entered twice, with a given twin once on the left and once on the right ("double-entry table"), so as to remove the ambiguity as to which variable a twin should be assigned.

This is illustrated with the 6-by-2 double-entry matrix in Table 2(a). The first three rows of this table are the same as in the deviation-score matrix. The last three rows are obtained by interchanging the first and second element in each row. As a result (and in contrast to the 3-by-2 deviation-score matrix), columns p and q both sum to zero, so that each column contains deviation scores. Hence, the variances of both columns p and q are simply the sum of squares divided by $2N-1$ ($= 5$ in the example), and the covariance is given by the sum of products (24) divided by 5. Since both variances are equal, the correlation is $r_l' = \text{covariance/variance} = (24/5)/(80/5) = .3$.

Intraclass correlations are measures of similarity within classes, in the present case, pairs of twins. They can also be computed, without any need for double-entry tables, from the mean squares of the 3-by-2 matrix of deviation scores (Harris, 1913):

$$r_l' := (SSB - SSW)/(SSB + SSW),$$

where SSB and SSW denote the between and within sums of squares, respectively. This is illustrated numerically in Part (c) of Table 2. The numerical value obtained from the sums of squares, .3, is the same as that obtained from the double-entry table.

For some time, it has been standard practice to define intraclass correlations in terms of means squares instead:

$$r_l := (MSB - MSW)/(MSB + MSW). \quad (1.5)$$

In this case, the value becomes $r_l = .472$. To distinguish this coefficient from r_l' obtained from the double-entry table (or, equivalently, from the sums-of-squares formula), the prime will be dropped. Although, in this small numerical example, both coefficients are different, their values will be very close for more realistic sample sizes.

Since r_l has become standard, it will be used from now on. For notational convenience, r will be retained for the population parameter, and the subscript l will be replaced with an indicator of the specific twin type. Thus, r_{MZT} stands for the "intraclass correlation of monozygotic twins raised together." On substituting the appropriate mean squares from Table 3 in (1.5), one obtains the correlations in Table 4. On dropping $\text{var}(d)$, one obtains the correlations under the more restrictive additive model (1.2). For example, Table 4 shows that, in terms of intraclass correlations, the re-

TABLE 4
Intraclass correlations under the dominance model (1.1)

	<i>Reared Together</i>	<i>Reared Apart</i>
MZs	$[\text{var}(a) + \text{var}(d) + \text{var}(e)]/\text{var}(y)$	$[\text{var}(a) + \text{var}(d)]/\text{var}(y)$
DZs	$[\text{.5 var}(a) + \text{.25var}(d) + \text{var}(e)]/\text{var}(y)$	$[\text{.5var}(a) + \text{.25var}(d)]/\text{var}(y)$

NOTE: Setting $\text{var}(d) = 0$ gives the intraclass correlations under (1.2).

liability r_{yy} is given by r_{MZT} under both models (1.1) and (1.2).

Heritability Indices. Since, by definition, a heritability index is supposed to indicate the proportion of genetic variance in the total systematic variance, the two models (1.1) and (1.2) give rise to two different heritability indices in common use:

$$\begin{aligned} &\text{broad heritability} \\ h_B &:= [\text{var}(a) + \text{var}(d)] / [\text{var}(a) + \text{var}(d) + \text{var}(e)] \end{aligned} \quad (1.6)$$

for the more comprehensive dominance model (1.1), and

$$\begin{aligned} &\text{narrow heritability} \\ h_N &:= \text{var}(a) / [\text{var}(a) + \text{var}(e)] \end{aligned} \quad (1.7)$$

for the more restrictive additive model (1.2).

Some authors use $\text{var}(y)$ in the denominator. Since, by Spearman-Brown, reliability varies with test length, such a definition renders heritability a function of the arbitrary test length.

Conventional Heritability Estimates. Over the years, numerous computational formulas have been advanced for estimating heritability from intra-class correlations. The underlying statistical rationale is usually obscure, if it is not lacking entirely. A case in point is Holzinger's h^2 , one of the oldest (e.g., Newman, Freeman, & Holzinger, 1937) and, until quite recently, one of the most widely used "heritability estimates":

$$\begin{aligned} h^2 &:= (r_{MZT} - r_{DZT}) / (1 - r_{DZT}) \\ &= \text{var}(a) / [\text{var}(a) + 2\text{var}(z)]. \end{aligned} \quad (1.8)$$

This index is meaningless because, as can be seen, it contains no environmental variance, $\text{var}(e)$, at all. Hottelling (in Newman, Freeman, & Holzinger, 1937), actually proceeded from the strictly additive model (1.2), but then made a mistake in his derivations of h^2 . As a result, if one equates K. Holzinger's coefficient (1.8) with narrow heritability (1.7), one arrives at the absurd conclusion that dizygotic twins share no genes (Schönemann, 1989). Similarly, A. R. JENSEN'S (1967) coefficient,

$$JHR = 2(r_{MZT} - r_{DZT}) = r_{MZT} h_N = r_{yy} h_N, \quad (1.9)$$

is problematic, since this coefficient varies with the length of the test which is arbitrary. If (1.6) is corrected for attenuation by dividing by $r_{yy} = r_{MZT}$ under

the additive model (1.2), then one arrives at Nichols's index,

$$\begin{aligned} HR &= 2(r_{MZT} - r_{DZT}) / r_{MZT} \\ &= \text{var}(a) / [\text{var}(a) + \text{var}(e)] \\ &=: h_N, \end{aligned} \quad (1.10)$$

which indeed coincides with narrow heritability, h_N , if (and only if) the strictly additive model (1.2) fits the data. On the other hand, under the dominance model (1.1), HR overestimates h_B :

$$\begin{aligned} HR &:= 2(r_{MZT} - r_{DZT}) / r_{MZT} \\ &= [\text{var}(a) + 1.5\text{var}(d)] / [\text{var}(a) + \text{var}(d) + \text{var}(e)] \\ &\geq [\text{var}(a) + \text{var}(d)] / [\text{var}(a) + \text{var}(d) + \text{var}(e)] \\ &=: h_B. \end{aligned}$$

From Table 3 one finds that under the dominance model (1.1) broad heritability is given, instead, by the ratio

$$\begin{aligned} r_{MZA} / r_{MZT} &= [\text{var}(a) + \text{var}(d)] / [\text{var}(a) \\ &\quad + \text{var}(d) + \text{var}(e)] \\ &=: h_B. \end{aligned} \quad (1.11)$$

PROBLEMS

Even the more comprehensive model (1.1) is, at best, only a rough approximation to reality. In particular, the omission of a gene-environment interaction variable and the assumption that twins raised apart are raised in completely uncorrelated environments have often been chided as patently implausible.

Regardless of how one assesses the intuitive plausibility of some of these simplifying assumptions, it is clear that if they are violated by the data, then the whole model becomes invalid and, with it, all heritability estimates derived from it. As D. Wahlsten (1990) noted, the underlying assumptions have rarely been tested with any stringency, if only because the conventional statistical tests lack power to detect violations if the sample sizes are small, as they often are in twin research. For global measures of fit, these difficulties compound, since global tests tend to conceal, rather than reveal, violations of critical assumptions. This was graphically demonstrated in two reanalyses of previously published data sets.

On reanalyzing the J. Shields (1962) MZT/MZA data, P. H. Schönemann (1990) found that two critical conditions implied by model (1.2) were violated in

seven out of eight cases, while a third was violated in eight out of eight cases. Many least-squares estimates of variances were negative, a telltale sign of trouble that an iterative fitting algorithm, designed to keep all estimates in the admissible range, would have masked. On refitting a purely environmental model, all variance estimates became nonnegative and the overall fit improved by a factor of 2. Similarly, P. H. Schönemann and R. D. Schönemann found numerous violations of model (1.2) in the T. Osborne (1980) MZT/DZT data. In particular, many heritability estimates computed according to Nichols's HR formula (1.10) exceeded unity, but Holzinger's nonsensical h^2 values appeared innocuous. Again, as with the Shields data, many variance estimates were negative under the additive model (1.2). On refitting these data with a qualitatively different, purely environmental model, most variance estimates became nonnegative, and the overall fit improved by a factor of 14.

A lesson to be learned from these reanalyses is that global tests of fit should be eschewed and, wherever possible, replaced with a series of narrower tests focused on specific conditions that the underlying models imply. For this to be possible, the model must first be spelled out in complete detail, including all its simplifying covariance assumptions. Such a more prudent research strategy does not seem to have won wide acceptance yet.

More generally, the above findings urge caution in interpreting unreasonably high heritability estimates of mental traits, especially of IQ, because statistical and conceptual errors that invalidate empirical inferences based on twin research went undetected for many years and, because in fair comparisons with alternative, purely environmental models, with the same number of parameters as models (1.1) and (1.2), the fit improved quantitatively and qualitatively, thereby invalidating previously reported heritability estimates derived from variants of the standard biometric model. Finally, one should always keep in mind that purely descriptive models, even if they do fit, are never conclusive, because one can never rule out the possibility that an entirely different model than the one used may fit the same data much better.

(See also: GENETICS, BEHAVIOR; TWIN STUDIES OF INTELLIGENCE.)

BIBLIOGRAPHY

- FISHER, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh*, 52, 399–433.
- GALTON, F. (1974). *Memories of my life*. New York: AMS Press. (Original work published 1908.)
- HARRIS, J. A. (1913). On the calculation of intra-class and inter-class coefficients of correlation from class moments when the number of possible combinations is large. *Biometrika*, 446–467.
- HIRSCH, J. (1981). To “unfrock the charlatans.” *Sage Relations Abstracts*, 6, 1–67.
- JENSEN, A. R. (1967). Estimation of the limits of heritability of traits by comparison of monozygotic and dizygotic twins. *Proceedings of the National Academy of Sciences*, 58, 149–159.
- KAMIN, L. (1974). *The science and politics of IQ*. Hillsdale, NJ: Erlbaum.
- KEMPTHORNE, O. (1978). Logical, epistemological, and statistical aspects of the nature–nurture data interpretation. *Biometrics*, 34, 1–23.
- NEWMAN, H., FREEMAN, F., & HOLZINGER, K. (1937). *Twins: A study of heredity and environment*. Chicago: University of Chicago Press.
- OSBORNE, T. (1980). *Twins: Black and white*. Athens, GA: Foundation for Human Understanding.
- PEDERSEN, N. L., PLOMIN, R., NESSELROADE, J. R., & MCCLEARN, G. E. (1992). A quantitative genetic analysis of cognitive abilities during the second half of the life span. *Psychological Science*, 3, 346–353.
- PLOMIN, R., & BERGEMAN, C. S. (1991). The nature of nurture: Genetic influence on “environmental” measures. *Behavioral and Brain Sciences*, 14, 373–385.
- ROUBERTOUX, P. L., & CAPRON, C. (1990). Are intelligence differences hereditarily transmitted? *Cahiers des Psychologie Cognitive/European Bulletin of Cognitive Psychology*, 10, 555–594.
- SCHÖNEMANN, P. H. (1989). New questions about old heritability estimates. *Bulletin of the Psychonomic Science*, 27, 175–178.
- SCHÖNEMANN, P. H. (1990). Environmental versus genetic variance component models for identical twins: A critique of Jinks and Fulker's reanalysis of the Shields data. *Cahiers de Psychologie Cognitive/European Bulletin of Cognitive Psychology*, 24, 734–745.

- SCHÖNEMANN, P. H., & SCHÖNEMANN, R. D. (1991). We wondered where the errors went. *Behavioral and Brain Sciences*, 14, 404–405.
- SCHÖNEMANN, P. H., & SCHÖNEMANN, R. D. (1994). Environmental versus genetic models for Osborne's personality data on identical and fraternal twins. *Cahiers de Psychologie Cognitive/European Bulletin of Cognitive Psychology*.
- SCOTT, J. P. (1987). Why does human twin research not produce results consistent with those from nonhuman animals? *Behavioral and Brain Sciences*, 10, 39.
- SHIELDS, J. (1962). *Monozygotic twins brought up apart and brought up together*. London: Oxford University Press.
- STEIGER, J. H., & SCHÖNEMANN, P. H. (1976). A history of factor indeterminacy. In S. Shye (Ed.), *Theory construction and data analysis in the behavioral sciences* (pp. 136–178). San Francisco: Jossey-Bass.
- WAHLSTEN, D. (1990). Insensitivity of the analysis of variance to heredity–environment interaction. *Behavioral and Brain Sciences*, 13, 109–161.
- WILSON, E. B. (1928). Review of “The abilities of man, their nature and measurement” by C. Spearman. *Science*, 67, 244–248.

PETER H. SCHÖNEMANN

HIERARCHICAL THEORIES OF INTELLIGENCE Some theories about the structure of intelligence emphasize one general ability (e.g., Humphreys, 1985; Jensen, 1987; & Spearman, 1927), while other theories emphasize several specialized abilities (e.g., Gardner, 1983; Guilford, 1967; Thurstone, 1938). The conflict between these types of theories is resolved in hierarchical theories, which involve both general and specialized abilities.

Two major categories of hierarchical theories may be distinguished. In one of these a hierarchy of dimensions of increasing generality is constructed, and in the other category the hierarchy is based on levels of functions.

HIERARCHICAL MODELS BASED ON GENERALITY OF DIMENSIONS

Factor analysis is a statistical technique through which performance on different tasks may be analyzed in terms of a limited number of underlying dimen-

sions, or factors, of ability. However, there may be relations among the factors that may be explained by another set of underlying abilities. Thus, a hierarchical model may be constructed through a factor analysis. It may, for example, be hypothesized that a single factor is sufficient to account for the intercorrelations among the factors. Such a factor is called a “second-order” factor. If a single factor cannot account for the correlations among the factors, one or more additional second-order factors may be introduced. Should there be several second-order factors, these may be correlated, and to account for these correlations a third-order factor may be introduced, and so on. Thus, with this approach a hierarchy of factors is built up, starting from below with a large number of narrow first-order factors and ending up at the top of the hierarchy with one, or a few, broad higher-order factors.

The most influential and well-known hierarchical model is the theory of fluid and crystallized ability developed by Raymond B. CATTELL and John Horn. The theory was first formulated by Cattell (1943), who argued that there is not one general factor of intelligence but two, which he labeled fluid and crystallized intelligence. Systematic empirical research was not reported until considerably later (e.g., Cattell, 1963; Horn, 1968; Horn & Cattell, 1966), however, when several broad, second-order abilities were identified in factor-analytic studies. The two dimensions of most central importance are fluid intelligence (*Gf*) and crystallized intelligence (*Gc*), and the whole theory is often referred to as *Gf-Gc* theory (see FLUID AND CRYSTALLIZED INTELLIGENCE). The *Gc* dimension is thought to reflect individual differences associated with systematic influences of acculturation; below this broad ability, verbal-conceptual abilities and abilities representing knowledge in different domains are subsumed. The *Gf* dimension is thought to reflect effects of biological and neurological factors and factors such as incidental learning. This broad dimension subsumes inductive and deductive reasoning abilities.

Several additional second-order factors have also been identified (e.g., Horn & Cattell, 1966; Horn & Stankov, 1982). Among these are general visualization (*Gv*), which subsumes abilities involved in tasks requiring manipulation and transformation of figural information; general speediness (*Gs*); general fluency

(*Gr*), identified as indicating retrieval from long-term storage; and short-term apprehension and retrieval (*Gsm*), also known as short-term memory.

On the basis of a reanalysis of virtually all correlation matrices collected in research on cognitive abilities up to 1990, John Carroll (1993) presented a hierarchical model of three degrees of generality that is referred to as the "Three-Stratum Model." At the highest level there is a single factor of general intelligence (*g*, also referred to as *G*), and at the next highest level some ten broad factors are identified. At the lowest level the model includes at least some sixty narrow factors. The broad factors in the Three-Stratum Model largely correspond to factors described in previous research (see, e.g., Horn, 1965), so Carroll's model largely coincides with previous results, except, of course, that it is more comprehensive than any previous factor-analytic account.

The first factor-analytic model was developed by Charles SPEARMAN (1904) in England. This model emphasized a single, general ability. Through contributions by Cyril BURT (1940) and others, however, a technique of FACTOR ANALYSIS was developed that first extracts the general factor and then extracts factors (called group factors) that successively get more and more narrow. This analytical technique thus produces results that easily fit a hierarchical structure, and in the British research, hierarchical models have played an important role at least since the 1940s.

Philip E. VERNON (1950) presented an integration of results achieved in several studies in the form of a model. At the top of his hierarchical model is a *g* factor. The model also includes two major group factors: the verbal-numerical-educational (*v:ed*) factor, and the practical-mechanical-spatial-physical (*k:m*) factor. Given a sufficient number of tests, these major group factors may be subdivided into several minor group factors. Thus, the *v:ed* factor subdivides into different scholastic factors such as *v* (verbal) and *n* (number) group factors. The *k:m* factor may be subdivided into minor group factors such as perceptual, spatial, and mechanical abilities. In contrast to the American research, however, in the British research relatively little emphasis has been placed on the minor group factors.

The American and British hierarchical models seem quite different, and it has, among other things, been

observed that the important *Gf* dimension does not have any counterpart in the British models. However, comparisons between the models using a modern form of factor analysis that allows specific hypotheses to be tested indicate that the differences may be more apparent than real (e.g., Gustafsson, 1984, 1988; Undheim, 1981). It has thus been found that the correlation between the second-order *Gf* factor and the third-order *g* factor is so close to unity that these factors must be considered identical. Thus, a reason why there is no *Gf* factor among the major group factors in the Vernon model may be that the *g* factor accounts for all the variance attributed to *Gf*.

HIERARCHICAL MODELS BASED ON LEVELS OF FUNCTIONS

The models considered so far attempt to identify ability dimensions of different degrees of generality and to clarify the relations among the dimensions. They do not necessarily involve any explicit psychological theory to account for the hierarchical arrangement of abilities. There are theories, however, that do specify a psychological basis for the hierarchical arrangement of abilities.

Burt (1949) contributed a hierarchical theory of the organization of the human mind based upon a model of levels of functions. He took a starting point in theories of a hierarchically organized mind based on evolutionary ideas. According to such theories, mental processes may be regarded as consisting of systems within systems, each process being classifiable according to degree of complexity into one of four levels. Processes at the lowest level are assumed to consist of simple sensations or movements. The next level includes processes of perception and coordinated movement. The third level is referred to as the associative level and involves memory and habit formation. The fourth and highest level involves the apprehension or application of relations.

Burt used this model to organize the findings of a comprehensive summary of factor-analytic results, and at each of the levels he reported evidence of group factors, which in some cases also split into subfactors. He interpreted the results as supportive of the hierarchical approach, even though the hierarchical struc-

ture as such could not be evaluated with these studies. The alternative hierarchical model proposed by Vernon (1950) came, however, to gain more popularity than the Burt model.

Another model with a classification of factors into a hierarchical scheme based upon levels of functions has been contributed by Horn (e.g., 1985, 1986, 1989). This model may be regarded as an extension of Burt's model, and it is, of course, supported by a much wider base of empirical findings than was available to Burt. The model organizes the abilities within an information-processing hierarchy with levels of sensory reception, association-processing, perceptual organization, and relation education.

The perceptual organization level, which provides input to the *Gf-Gc* processes, includes *Gv* and a factor of general auditory (*Ga*) competence, reflecting capacities for dealing with the complexities of sound. This level also includes the *Gs* dimension. The level of association processing includes two dimensions representing memory capacities. One of these, short-term acquisition and retrieval (SAR), reflects the capacity to store and retrieve information over such short periods of time as a minute or two, and the other, tertiary storage and retrieval (TSR), identifies an ability to retrieve information stored a considerable time before the measurement. At the lowest level—that of sensory reception—one factor represents the acuity of visual sensory detectors (vSD), and another represents the acuity of auditory sensory detectors (aSD).

Another development of the *Gf-Gc* theory, which partly relies on neuropsychological concepts, has been presented by Cattell (1987). The so-called triadic theory includes abilities of three different kinds. "General capacities" (e.g., *Gf*, *Gs*, and *Gr*) represent limits to brain action as a whole. Another class of abilities is referred to as "provincial powers." These correspond to sensory area factors, such as visualization, auditory structuring ability, and motor and kinesthetic abilities. The third class of abilities is referred to as "agencies," which are narrow abilities to perform in different areas of cultural content. According to the triadic theory, the three kinds of abilities combine in joint action in observed behavior (Cattell, 1987, p. 366).

Both Horn's and Cattell's hierarchical theories are complex and complete theoretical systems, which are not easily subjected to empirical tests. Such theoretical

systems are best evaluated against criteria of fruitfulness, however, and it remains to be seen how successful these and other hierarchical theories are as guides of further research.

CONCLUSION

Hierarchical theories have been available for a long time, but they gained increased popularity during the 1980s and early 1990s (Lohman, 1989). One reason for this is that more complete and useful models and theories had become available. Development was still continuing in the early 1990s.

Even though the question of whether a dimension of general cognitive ability should be recognized remains somewhat controversial (e.g., Horn, 1989), many theorists now agree that at least three categories of ability dimensions should be recognized: general cognitive ability, broad abilities, and narrow abilities. Much further research is needed, however, to clarify the nature of these ability dimensions, their relationships, and their psychological interpretation. One particularly interesting and urgent task for further research is to achieve a better integration between the descriptive, factor-analytic, hierarchical models, on the one hand, and the process-oriented hierarchical theories, on the other.

BIBLIOGRAPHY

- BURT, C. (1940). *The factors of the mind: An introduction to factor-analysis in psychology*. London: University of London Press.
- BURT, C. (1949). The structure of the mind: A review of the results of factor analysis. *British Journal of Educational Psychology*, 19, 100–111, 176–199.
- CARROLL, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York: Cambridge University Press.
- CATTELL, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin*, 40, 153–193.
- CATTELL, R. B. (1963). Theory of fluid and crystallized intelligence: A critical experiment. *Journal of Educational Psychology*, 54, 1–22.
- CATTELL, R. B. (1987). *Intelligence: Its structure, growth and action*. New York: North-Holland.
- GARDNER, H. (1983). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

- GUILFORD, J. P. (1967). *The nature of human intelligence*. New York: McGraw-Hill.
- GUSTAFSSON, J. E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence*, 8, 179–203.
- GUSTAFSSON, J. E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 35–71). Hillsdale, NJ: Erlbaum.
- HORN, J. L. (1968). Organization of abilities and the development of intelligence. *Psychological Review*, 72, 242–259.
- HORN, J. L. (1985). Remodeling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 267–300). New York: Wiley.
- HORN, J. L. (1986). Intellectual ability concepts. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 3, pp. 35–77). Hillsdale, NJ: Erlbaum.
- HORN, J. L. (1989). Models of intelligence. In R. L. Linn (Ed.), *Intelligence: Measurement, theory and public policy* (pp. 29–73). Urbana: University of Illinois Press.
- HORN, J. L., & CATTELL, R. B. (1966). Refinement and test of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology*, 57, 253–270.
- HORN, J. L., & STANKOV, L. (1982). Auditory and visual factors of intelligence. *Intelligence*, 6, 165–185.
- HUMPHREYS, L. G. (1985). General intelligence: An integration of factor, test and simplex theory. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 201–224). New York: Wiley.
- JENSEN, A. R. (1987). The g beyond factor analysis. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.), *Buros-Nebraska Symposium on Measurement and Testing: Vol. 3. The influence of cognitive psychology on testing* (pp. 87–142). Hillsdale, NJ: Erlbaum.
- LOHMAN, D. F. (1989). Theory and research in intelligence. *Review of Educational Research*, 59(4), 333–373.
- SPEARMAN, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology*, 15, 201–293.
- SPEARMAN, C. (1927). *The abilities of man*. London: Macmillan.
- THURSTONE, L. L. (1938). Primary mental abilities. *Psychometric Monographs*, No. 1.
- UNDHEIM, J. O. (1981). On intelligence IV: Toward a restoration of general intelligence. *Scandinavian Journal of Psychology*, 22, 251–265.
- VERNON, P. E. (1950). *The structure of human abilities*. London: Methuen.

JAN-ERIC GUSTAFSSON

HISPANICS Hispanics are the largest linguistic minority group in the United States. This racial/ethnic group includes persons of Mexican origin (Chicanos or Mexican Americans), Puerto Ricans, Cubans, and persons of Central or South American origin. By one estimate the Hispanic population will reach 30 million in the United States by the year 2010 (Cervantes & Acosta, 1992). Because of the diversity of the Hispanic ethnic group, Hispanic Americans do not comprise a single and cohesive cultural group.

Current demographic indicators reflect the problems faced by Hispanic Americans in the United States. L. Baruth and M. Manning (1992) cite information from the 1990 United States census indicating that Hispanic Americans have the highest school dropout rates of any ethnic group. In addition, a substantial proportion (39%) of Hispanic Americans under the age of 16 lived below the poverty level in 1989. D. McShane and V. Cook (1985) also report that Hispanic children show extremely high rates of particular medical problems such as middle ear disease. Thus, the concerns confronting the subgroups comprising this racial/ethnic segment of the U.S. population are considerable.

HISPANIC INTELLIGENCE AS MEASURED BY STANDARDIZED INSTRUMENTS

Numerous studies have addressed the intelligence of Hispanic Americans. Consistent throughout the majority of studies is the finding that Hispanics tend to score lower than whites on many IQ measures. Various factors may contribute to this lower-than-average IQ including number of years in U.S. schools, socioeconomic status, environment, and language dominance (English or Spanish).

One of the most comprehensive research projects was conducted by Thomas Sowell (1978) for the Urban Institute. Sowell examined approximately 70,000 IQ records of students from various ethnic groups

across the United States that had been recorded over a period of fifty years. The overall goal was to determine if there were historical patterns of mental test scores. According to Sowell's report, the average (median) IQs of Mexican Americans ranged from 83 to 87 between 1940 and 1970; Puerto Rican median IQs ranged from 79 to 84 during the same period. Findings also indicated that the average IQ for Puerto Ricans increased depending upon the number of years spent in the U.S. schools. Thus, students with one to two years of U.S. schooling had an average IQ of 72. Students with nine to ten years of U.S. schooling had an average IQ of 93. These results indicate that currently the IQs of Hispanic Americans tend to be below the average IQ of whites as obtained on standardized tests and that this difference may decrease as a function of more time in the United States and exposure to the American school system.

In their reviews and in agreement with Sowell (1978), D. McShane and V. Cook (1985) and R. Valencia (1979) also report that Hispanic children tend to score lower than their Caucasian counterparts on standardized IQ measures. As with white children, the sociocultural background of these Hispanic children appears to have a different impact on IQ scores; that is, Hispanic children of middle socioeconomic status and from urban environments tend to score higher than those Hispanic children from poorer and more rural backgrounds. These authors also note that because of their relatively lower scores, Hispanic children may be at risk to be identified as mentally retarded.

K. Gerken (1978) has pointed out that researchers cannot assume that the low scores obtained by Mexican-American children are indicative of less highly developed intellectual ability. She notes that many important variables such as linguistic background are often not addressed in studies. Gerken specifically examined the role of language in IQ by comparing scores obtained by these groups of children: bilingual, Spanish-dominant, and English-dominant Mexican-American children. The measures used included: (1) a nonverbal scale of ability that utilized nonverbal examiner directions (Leiter International Performance Scale; Leiter, 1969); (2) a nonverbal scale with English verbal directions (Wechsler Preschool and Primary

Scale of Intelligence [WPPSI] Performance Scale; Wechsler, 1967); and (3) a verbal scale with English verbal directions (WPPSI Verbal Scale; Wechsler, 1967). The Spanish-dominant children obtained an overall IQ of 62 on the WPPSI Verbal Scale, 88 on the WPPSI Performance Scale, and 90 on the Leiter. Bilingual children obtained an overall IQ of 84 on the WPPSI Verbal Scale, 104 on the WPPSI Performance Scale, and 109 on the Leiter. English-dominant children obtained an overall IQ of 104 on the WPPSI Verbal Scale, 109 on the WPPSI Performance Scale, and 109 on the Leiter. Results indicate that Mexican-American children scored higher on the nonverbal scale with nonverbal directions and the nonverbal scale with verbal directions in comparison with the verbal scale with verbal directions. In addition, the dominant language of these children also related to IQ as Spanish-dominant children obtained the lowest IQ scores, particularly on the tests with English verbal directions.

Support for this finding is also noted in a review of the literature by R. Valencia, R. Henderson, and R. Rankin (1981). They report that performance on intelligence measures for Mexican-American children is closely linked to their language status. In general, Mexican-American children with English as their native language had higher scores on intelligence measures in comparison with their Spanish-dominant peers.

Gerken (1978) reports that although some researchers have used nonverbal measures of ability in the hope of eliminating the role of language, nonverbal tests may measure different abilities than those assessed by verbal tests. Also, nonverbal tests often include a verbal component such as verbal instructions. Gerken also acknowledges that nonverbal tests do not correlate well with school achievement. Therefore, obtaining higher scores on these measures may be of little practical importance from an educational point of view.

USAGE OF STANDARDIZED MEASURES WITH HISPANICS

Concerns regarding the usage of cognitive measures with various racial/ethnic groups, including Hispanic, has been reported frequently in the literature. The

usage of standardized IQ measures in special educational placement decisions (e.g., the determination that a child is mentally retarded or is learning disabled) of minority children has been especially controversial. In particular, researchers often cite issues pertaining to the reliability and validity of IQ tests across racial/ethnic groups. **RELIABILITY** refers to the record of the test in measuring intelligence in a reproducible and consistent fashion. Would an individual examined more than once with the same test achieve essentially the same score during each administration? **VALIDITY** refers to the extent to which a test measures the construct (intelligence) that it purports to assess. These concepts are of particular relevance with respect to usage of various intellectual instruments with Hispanics.

In general, researchers have found that intelligence tests like the Wechsler scales yield consistent and reliable scores for Hispanic groups (Dean, 1977; McShane & Cook, 1985). Studies also indicate that standardized instruments assess the same dimensions of intelligence for Caucasian and Mexican-American children (Geary & Whitworth, 1988; Gutkin & Reynolds, 1980). Although the structure of cognitive abilities (comprising overall intelligence) does not appear to differ between racial/ethnic groups, the level of abilities (higher or lower IQ scores) may vary (Geary & Whitworth, 1988). Differences exist between the overall IQ scores obtained for Hispanics and other groups. For example, Hispanic children generally score lower than Caucasians on IQ tests but higher overall than black children (McShane & Cook, 1985). Some argue that these discrepancies are indicative of test bias.

Attempts have been made to address the "bias" issue through adjustment of scores obtained on intelligence measures based on social, cultural, and linguistic variables. One such attempt was the development of the **SYSTEM OF MULTICULTURAL PLURALISTIC ASSESSMENT (SOMPA; Mercer, 1979)**. The SOMPA yields an Estimated Learning Potential (ELP) based upon a comparison of an individual's scores with other people of the same sociocultural background. Although controlling for cultural variables related to the measurement of intelligence, the ELP does not appear to predict or correlate with academic achievement for minority children as accurately as other standardized IQ scores (Figueroa & Sassenrath, 1989).

PATTERNS OF INTELLECTUAL ABILITIES FOR HISPANIC AMERICANS

Drawing definitive conclusions on the ability patterns of Hispanic children has been difficult given the number of factors that may have an impact on ability patterns. For example, McShane and Cook (1985) reviewed over seventy empirical studies on the performance of Hispanics on the **WECHSLER SCALES OF INTELLIGENCE**, the most widely used individualized intelligence tests in the United States. These authors report that factors such as age, socioeconomic status, sex, sociocultural background, and urban-rural residence significantly influenced the IQ scores obtained by the Hispanic racial/ethnic subgroups studied. Higher visual-spatial abilities in comparison with verbal abilities are frequently cited for Hispanics as noted in the following discussion.

The extensive literature review conducted by McShane and Cook (1985) reported that Hispanic Americans may demonstrate strengths in visual-spatial abilities as measured by performance subtests in comparison with verbal abilities measured by the verbal subtests. This finding is fairly consistent across the age span when assessed by the Wechsler scales. The discrepancy between the higher visual-spatial abilities (i.e., nonverbal) relative to verbal abilities is approximately 20 points on the Wechsler Preschool and Primary Scale of Intelligence (WPPSI; Wechsler, 1967) and, comparably, 10 to 15 points on the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974).

In common with the studies cited earlier, the results of a study by L. Laosa (1984) indicate that Mexican children (aged 2 years, 6 months) score lower in ability than their Caucasian peers on tasks involving verbal, quantitative, and short-term memory. The abilities tapped by these tests are often linked to success in school. Thus, Mexican students not surprisingly generally perform in school at achievement levels lower than Caucasians. Using these results, Laosa suggests that Mexicans appear to be disadvantaged in measured abilities that may be required for success in school.

R. Taylor and S. Richards (1991) compared the abilities on subtests of the Wechsler scales of black,

Hispanic, and Caucasian children when overall IQ was held constant (i.e., all groups were equated on IQ). These authors found that Hispanic children performed better on nonverbal than verbal abilities. Specifically, they scored highest on the Picture Completion, Block Design, and Object Assembly subtests of the Wechsler scale (see ETHNICITY, RACE, AND MEASURED INTELLIGENCE for a description of these measures). These three subtests are believed to assess visual-spatial abilities.

D. Saccuzzo, N. Johnson, and G. Russell (1992), however, report that the discrepancy between verbal and these nonverbal, visual-spatial abilities noted in previous literature does not hold true for gifted Hispanic children. In their comparative study of gifted Caucasian, Filipino, and Hispanic children using the Wechsler Intelligence Scale for Children-Revised, gifted Hispanic children obtained higher verbal than performance scores. The difference, however, was small (Verbal 130, Performance 128) and of no practical significance for the Hispanic group. Parenthetically, although gifted African-American and Caucasian children also obtained higher verbal than performance scores (African-American: Verbal 130, Performance 123; Caucasian: Verbal 135, Performance 130), the Filipino group did not (Verbal 126, Performance 128). Regarding the Hispanic group, Saccuzzo, Johnson, and Russell (1992) note that their findings with gifted children are different from those obtained from children with average and below-average IQs. Specifically, the latter found that as the IQ for Hispanic children increases, the discrepancy between verbal and performance abilities decreases.

HYPOTHESES RELATING DIFFERENCES IN INTELLIGENCE TO FIELD DEPENDENCE/FIELD INDEPENDENCE

Explanations regarding IQ differences have focused on issues pertaining to cognitive style, particularly field dependence and field independence. As a cognitive style, *field dependence* refers to tendencies to attend more to cues in the environment, and *field independence* pertains to a focus on internalized cues (Witkin, 1974). Many researchers have noted that overall, Hispanic children tend to be more field dependent when com-

pared to their Caucasian peers who are more field independent (Kagan & Zahn, 1975; Ramirez & Price-Williams, 1974; Sanders, Scholz & Kagan, 1976). The relationship between field dependence/field independence and various test scores may vary, however, for Hispanic children based upon acculturation issues and environmental setting (Knight et al., 1978).

In relation to measures of intelligence, field-dependent children tend to score lower overall than field-independent children (Ramirez & Castaneda, 1974). This situation is consistent with findings that Hispanics generally score lower on IQ tests in comparison with their Caucasian peers who are more field independent. M. Ramirez and A. Castaneda explain this finding as a consequence of field-independent individuals' relatively higher scores on tests assessing analytical abilities. Thus, the higher scores obtained by field-independent individuals may derive from their relative strengths in analytic abilities. An understanding of field dependence/field independence in relation to cultural factors and hemispheric differences is helpful in examining the intelligence of the Hispanic racial/ethnic group.

Cultural Factors and Field Dependence/Field Independence. Cultural factors affect the development of cognitive style and performance on IQ measures. Differences in traditional versus dualistic communities and generational membership (i.e., generation in the United States) have been specifically reported in the literature. Traditional communities are those that are rural and tend to have closer ties with Mexico (Ramirez & Castaneda, 1974). Mexicans are usually the majority of the population in traditional communities that are located near the United States-Mexico border in South Texas, New Mexico, and California. Dualistic communities tend to be located in more metropolitan areas. Researchers have indicated that more than families in dualistic communities, traditional communities appear to emphasize conformity to authority, adherence to conventional practices, and a priority to family and community over individual needs (Buriel, 1975; Ramirez & Castaneda, 1974; Laosa & DeAvila, 1979). Mexicans growing up in dualistic communities generally experience greater pressure to integrate the values of the mainstream American culture, although evidence exists for dualistic communities experiencing pressure to uphold traditional Mexican values as well.

As one might expect from these characteristics, those Hispanics who were exposed to a dualistic environment tended to have a relatively field-independent cognitive style, whereas children in the traditional communities tended to be more field dependent. In addition, over time, both types of communities demonstrated a progressive increase in field independence (Laosa & DeAvila, 1979).

Changes in the relationship between cognitive style within the traditional community are also noted in examination of the "intergenerational cognitive styles" of Mexican-American children (Buriel, 1975). Findings indicate that there are generational differences in cognitive style. In one study, first- and third-generation Mexican-American children demonstrated greater field dependence than second-generation children who scored closer to the Caucasian norm of greater field independence. The study also indicated that third-generation subjects demonstrated the highest level of field dependence.

HEMISPHERIC DIFFERENCES AND FIELD DEPENDENCE/FIELD INDEPENDENCE

Ramirez and Castaneda (1974) relate field dependence/field independence to differences in the processing of information in the two hemispheres of the brain. This relationship is somewhat oversimplified given the earlier discussion of these constructs and the impact of community and generation.

Characteristics of the field sensitive (field-dependent) cognitive style are analogous to those identified with the functioning of the right cerebral hemisphere, whereas characteristics of the field-independent cognitive style are similar to those identified as functions of the left hemisphere [p. 74].

The left hemisphere is linked to verbal abilities, and the right hemisphere is linked to visual-spatial abilities. Thus, findings of higher visual-spatial abilities (i.e., higher scores on the performance subtests of the Wechsler scale) in comparison with verbal abilities indicate that Hispanics may have strengths in right-hemisphere abilities. Studies have indicated that Hispanic children obtain high right-hemisphere scores (Mitchell, 1987-1988). This hemispheric difference

may decrease as the overall intelligence of the individual increases as in the case of gifted students.

LIMITATIONS OF RESEARCH

Studies examining the intelligence scores obtained by Hispanic groups are often criticized in light of several limitations that affect findings. Valencia (1979) notes in his review that the research on Mexican intelligence in the second half of the twentieth century has failed to control for the following concerns: socioeconomic status of the participants, language and test translation issues, and cultural loading (the cultural content of IQ tests linked to Caucasian middle-class culture). The following discussion highlights limitations related to these issues.

Socioeconomic Status. A major finding in the literature has been that, for each racial/ethnic group studied, a relationship exists between socioeconomic status and overall IQ. When higher and lower socioeconomic groups are compared, differences in the level (higher or lower scores) of abilities often occur, but the overall pattern (profile of abilities across subtests) remains similar (Lesser, Fifer, & Clark, 1965). Thus, Hispanics with higher socioeconomic status tend to score higher on intelligence tests than Hispanics with lower socioeconomic status (McShane & Cook, 1985). This finding is consistent throughout much of the Hispanic literature and is consistent with the findings in numerous studies of Caucasian and Asian individuals.

Language and Test Translation Issues. The relationship between language and intellectual performance is unclear (Valencia & Rankin, 1985). For example, Olmedo (1981) found that tests written in Spanish are often mistakenly used with people who speak significantly different Spanish dialects. Difficulties in assessing the abilities of bilingual children also occur. For example, as discussed above Mexican-American children who are bilingual may demonstrate different patterns of abilities given the nature of "bilinguality" (Matluck & Mace, 1973, p. 371).

There are also problems in test translation. R. Samuda (1975) reports that translating tests into Spanish does not solve problems because the test content in the English version is culture-bound to the American culture. Thus, items restated in different languages may not overlap in meaning. In addition, when an En-

glish word is translated into Spanish, the Spanish word often has more syllables than the English word. Thus, these longer items (with more syllables) may reduce recall on measures of memory for Spanish speakers (Olmedo, 1981).

Valencia, Henderson, and Rankin (1981) further emphasize the importance of language in the assessment of intelligence. They state that the "single most powerful predictor of mental performance was a language/schooling factor consisting of the language spoken in the home, the language in which the test was administered, the level of educational attainment of the mother and father, and the country where the parents were educated" (p. 528). The most "competent children" in the study came from homes in which the dominant spoken language was English. In addition, these children were tested in English and their parents had attained higher levels of education in the United States rather than in Mexico. Valencia and associates suggest that parents who have been educated in the United States are more likely to communicate to their children the culture of the school environment than parents educated in Mexico. Many of the skills reinforced in the school environment are those that are tapped by tests of intelligence.

Cultural Loading. Because of the differences in the average IQs evidenced in the comparison between the Hispanic racial/ethnic group and their Caucasian peers, some researchers have suggested that cultural loading may have an impact on intelligence test scores. R. Figueroa (1983) indicates that cultural bias may be a function of cultural loading, that is, the degree to which the cultural background of the children being tested approximates the background of children in the samples on which the intelligence test was developed and normed. The more discrepant a child's culture is from the culture of the children upon which the test is based, the more culturally loaded a test may be and the more bias may be operating to lower IQ scores artificially.

CONCLUSIONS

In general, most studies have indicated that Hispanic Americans tend to have lower mean IQ scores than do Caucasians. The diversity of the subgroups that comprise this racial/ethnic classification, language

factors, cognitive style, socioeconomic status, cultural loading, and generational differences each make difficult the interpretation of the overall intelligence and ability patterns of Hispanic Americans. Therefore, the results of studies published to date must be interpreted with caution because of the potential role these as yet not well controlled variables potentially play in the measured intelligence of groups of Hispanic Americans.

(See also: ETHNICITY, RACE, AND MEASURED INTELLIGENCE.)

BIBLIOGRAPHY

- BARUTH, L. G., & MANNING, M. L. (1992). Understanding and counseling Hispanic American children. *Elementary School Guidance and Counseling*, 27, 113-122.
- BURIEL, R. (1975). Cognitive styles among three generations of Mexican American children. *Journal of Cross-Cultural Psychology*, 6(4), 417-429.
- CERVANTES, R. C., & ACOSTA, F. X. (1992). Psychological testing for Hispanic Americans. *Applied and Preventive Psychology*, 1, 209-219.
- DEAN, R. S. (1977). Reliability of the WISC-R with Mexican American children. *Journal of School Psychology*, 15(3), 267-268.
- FIGUEROA, R. A. (1983). Test bias and Hispanic children. *Journal of Special Education*, 17(4), 431-440.
- FIGUEROA, R. A., & SASSENATH, J. M. (1989). A longitudinal study of the predictive validity of the System of Multicultural Pluralistic Assessment (SOMPA). *Psychology in the Schools*, 26, 5-19.
- GEARY, D. C., & WHITWORTH, R. H. (1988). Is the factor structure of the WISC-R different for Anglo- and Mexican-American children? *Journal of Psychoeducational Assessment*, 6, 253-260.
- GERKEN, K. C. (1978). Performance of Mexican American children on intelligence tests. *Exceptional Children*, 44(6), 438-443.
- GUTKIN, T. B., & REYNOLDS, C. R. (1980). Factorial similarity of the WISC-R for Anglos and Chicanos referred for psychological services. *Journal of School Psychology*, 18, 34-39.
- KAGAN, S., & ZAHN, G. L. (1975). Field dependence and the school achievement gap between Anglo American and Mexican American children. *Journal of Educational Psychology*, 67(5), 543-650.

- KNIGHT, G. P., KAGAN, S., NELSON, W., & GUMBINER, J. (1978). Acculturation of second–third-generation Mexican American children: Field independence, locus of control, self-esteem, and school achievement. *Journal of Cross-Cultural Psychology*, 9(1), 87–97.
- LAOSA, L. M. (1984). Ethnic, socioeconomic, and home language influences upon early performance on measures of abilities. *Journal of Educational Psychology*, 76, 1178–1198.
- LAOSA, L. M., & DEAVILA, E. A. (1979). Development of cognitive styles among Chicanos in traditional and dualistic communities. *International Journal of Psychology*, 14, 91–98.
- LEITER, R. G. (1969). *General instructions for the Leiter International Performance Scale*. Chicago: Stoelting.
- LESSER, G. S., FIFER, G., & CLARK, D. H. (1965). Mental abilities of children from different social-class and cultural groups. *Monographs of the Society for Research in Child Development*, 30, (4, Whole No. 102).
- MATLUCK, J. H., & MACE, B. J. (1973). Language characteristics of Mexican-American children: Implications for assessment. *Journal of School Psychology*, 11, 365–386.
- MC SHANE, D., & COOK, V. (1985). Transcultural intellectual assessment: Performance by Hispanics on the Wechsler scales. In B. Wolman (Ed.), *Handbook of intelligence: Theories, measurement and application* (pp. 737–785). New York: Wiley.
- MERCER, J. R. (1979). *System of Multicultural Pluralistic Assessment (SOMPA) technical manual*. San Antonio, TX: The Psychological Corporation.
- MITCHELL, B. M. (1987–1988). Hemisphericity and creativity: A look at the relationships among elementary-age low-income Hispanic children. *Educational Research Quarterly*, 12(1) 2–5.
- OLMEDO, E. L. (1981). Testing linguistic minorities. *American Psychologist*, 36(10), 1078–1085.
- RAMIREZ, M., & CASTANEDA, A. (1974). *Cultural democracy, bi-cognitive development and education*. New York: Academic Press.
- RAMIREZ, M., & PRICE-WILLIAMS, D. R. (1974). Cognitive styles of children in three ethnic groups in the United States. *Journal of Cross-Cultural Psychology*, 5(2), 212–219.
- SACCUZZO, D. P., JOHNSON, N. E., & RUSSELL, G. (1992). Verbal versus performance IQs for gifted African-American, Caucasian, Filipino, and Hispanic children. *Psychological Assessment*, 4(2), 239–244.
- SAMUDA, R. J. (1975). *Psychological testing of American minorities: Issues and consequences*. New York: Harper & Row.
- SANDERS, M., SCHOLZ, J. P., & KAGAN, S. (1976). Three social motives and field independence-dependence in Anglo American and Mexican American children. *Journal for Cross-Cultural Psychology*, 7(4), 451–461.
- SOWELL, T. (1978). Race and IQ reconsidered. In T. Sowell (Ed.), *Essays and data on American ethnic groups* (pp. 203–238). Washington, DC: The Urban Institute.
- TAYLOR, R. L., & RICHARDS, S. B. (1991). Patterns of intellectual differences of black, Hispanic, and white children. *Psychology in the Schools*, 28, 5–8.
- VALENCIA, R. R. (1979). Comparison of intellectual performance of Chicano and Anglo third-grade boys on the Raven's Coloured Progressive Matrices. *Psychology in the Schools*, 16(3), 448–453.
- VALENCIA, R. R. (1983). Stability of the McCarthy Scales of Children's Abilities over a one-year period for Mexican-American children. *Psychology in the Schools*, 20, 29–34.
- VALENCIA, R. R., HENDERSON, R. W., & RANKIN, R. J. (1981). Relationship of family constellation and schooling to intellectual performance of Mexican American children. *Journal of Educational Psychology*, 73(4), 524–532.
- VALENCIA, R. R., & RANKIN, R. J. (1985). Evidence of content bias on the McCarthy Scales with Mexican American children: Implications for test translation and nonbiased assessment. *Journal of Educational Psychology*, 77(2), 197–207.
- WECHSLER, D. (1967). *Manual for the Wechsler Preschool and Primary Scale of Intelligence*. San Antonio, TX: The Psychological Corporation.
- WECHSLER, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised*. San Antonio, TX: The Psychological Corporation.
- WITKIN, H. A. (1974). Cognitive style across cultures. In J. W. Berry & P. R. Dasen (Eds.), *Cultural and cognition: Readings in cross-cultural psychology* (pp. 99–118). London: Methuen. (Reprinted from *International Journal of Psychology*, 1967, 2(4), 233–250.)

LISA A. SUZUKI
TERRY B. GUTKIN

HOLLINGWORTH, LETA S. (1886–1939)

The educational and adolescent psychologist Leta Anna Stetter Hollingworth was born May 25, 1886, “in a dugout on the White River, five or six miles from the site of what is now Chadron, Nebraska . . .” (H. L. Hollingworth, 1943, p. 2) in the even-now-rural

northwestern part of that state. Her mother, born Margaret Elinor Danley in Illinois in 1862, was “the daughter of . . . a migrant farmer and his wife” (p. 2). Her father, John G. Stetter, born in Virginia about 1855, came from a “family . . . of German-speaking stock” (p. 10). He held a variety of positions and initiated several enterprises. “Always garrulous and friendly, free with his resources to those for whom he had no responsibility, taking in fact all responsibility lightly, he became during his early and long residence on the frontier a well-known character in that corner of the state, and his name always provoked a tolerant smile of recognition. He was in fact always a boy, and never came to feel the burden of the world, as did his daughter, after him” (p. 11.) From such humble, seemingly improbable origins arose one of the great women of the first half of the twentieth century, working in the nation’s largest city.

She received her B.A. degree in 1906, graduating Phi Beta Kappa and summa cum laude from the University of Nebraska, Lincoln. In 1913 she received an M.A. degree and Master’s Diploma in Education from Teachers College, Columbia University, where she also earned her doctorate in 1916. The University of Nebraska awarded her an honorary LL.D. degree in 1938.

Leta Hollingworth began as a scientific feminist, one of the earliest and most experimentally rigorous. She strove to destroy such myths as the notion that women are incapacitated mentally and physically during their menstrual cycles and that the variability between men and women and the greater ability of men in most intellectual areas is so great that few women can aspire to achieve as well as the ablest men. At first, also, she was interested in low-IQ children. Then, in 1916 she tested an 8-year-old boy with Lewis M. Terman’s (1916) newly published Stanford-Binet Intelligence Scale, finding his IQ to be at least 187. From shortly thereafter, her research, development, service, and publications mainly concerned high-IQ boys and girls. She was on the faculty of Teachers College of Columbia University in New York City from 1916, starting as an instructor and rising to full professor, until her untimely death from stomach cancer in 1939 at age 53. In those relatively few years she revolutionized thinking about educational facilitation of the mentally gifted.

Also, she pioneered in administering to such youths difficult cognitive and physical tests designed for older students. She showed that, especially on the former, extremely high-IQ children could often equal or surpass Columbia University graduate students or Yale Law School entrants. This validation of above-age, above-grade-level testing is the foundation for the extensive present use with bright 12-year-olds throughout the United States of such difficult college-admission examinations as the Scholastic Aptitude Test (Stanley, 1990).

While Lewis M. TERMAN, far from New York City, was studying high-IQ youths as they grew up, Hollingworth was concentrating on helping them find the educational opportunities she was convinced they greatly needed and richly deserved. She expended much time and energy on creating special classes for them in New York City public schools. Concurrently, she experimented with various aspects of “children who test at or above 135 I.Q. (Stanford-Binet)”: size and strength, musical sensitivity, tapping rate, neuromuscular capacity, and the intelligence quotients of their siblings (Hollingworth, 1990, pp. 160–161).

A number of Hollingworth’s articles about intelligence and exceptionally intelligent children were published in magazines intended for the intelligent layperson. She was both scientific researcher and expositor, always eager to point out the advantages and responsibilities that great mental ability confers on its possessors, their parents and teachers, and society itself.

She was a prolific writer: nine books, eighty-two articles, many reviews, summaries, memoranda, and reports, and more than a thousand letters a year. Her best-known works are *Gifted Children* (1926) and *Children above 180 I.Q. Stanford-Binet* (1942). The latter was completed and published after her death by her husband, Harry L. Hollingworth, himself an outstanding psychologist. A year later, he published a loving but objective and factual biography of her, containing a complete record of her publications and of publications about her (Hollingworth, 1990, pp. 159–168). The biography, long out of print, was updated and republished in 1990.

Leta Hollingworth was one of the leading psychologists of her time and among the few women in those days who became full professors at major universities. She worked largely without grants, adequate office

space, or much secretarial help. Her work, virtually forgotten after her death, has been revived recently as a stellar exemplar of what professional women can do. She was the cofounder of the worldwide gifted-child movement, contemporary with Lewis M. Terman but working largely independently of him. Many of her insights into the meaning of high IQs and of the potentialities of extremely bright boys and girls are as valid today as when she produced her path-breaking *Gifted Children* in 1926. Although Terman provided a principal tool (his 1916 intelligence scale) needed to identify high-IQ youths, Hollingworth developed procedures for nurturing them educationally. Especially, she studied the academic concomitants and possibilities of high intelligence.

BIBLIOGRAPHY

- HOLLINGWORTH, H. L. (1990). *Leta Stetter Hollingworth: A Biography*. Bolton, MA: Anker. (Republication, with added foreword and bibliographies, of title issued originally in 1943 by the University of Nebraska Press, Lincoln).
- HOLLINGWORTH, L. S. (1914). Functional periodicity. *Contributions to Education*, 69. New York: Columbia University Teachers College Bureau of Publications.
- HOLLINGWORTH, L. S. (1914). Variability as related to sex differences in achievement. *American Journal of Sociology*, 19, 510–530.
- HOLLINGWORTH, L. S. (1926). *Gifted Children*. New York: Macmillan.
- HOLLINGWORTH, L. S. (1942). *Children Above 180 I.Q. Stanford-Binet*. Yonkers, NY: World Book.
- ROEPER REVIEW. (1990). 12(3, March), pp. 131–237. (Entire issue, except p. 238, was devoted to “Leta Stetter Hollingworth, 1886–1939”).
- STANLEY, J. C. (1990). Leta Hollingworth’s contributions to above-level testing of the gifted. *Roeper Review*, 12, 166–171.
- TERMAN, L. M. (1916). *The Measurement of Intelligence*. Boston: Houghton Mifflin.

JULIAN C. STANLEY

HUMPHREYS, LLOYD G. (1913–) Lloyd Girton Humphreys was born in Lorane, Oregon, to a working-class family. He earned a B.S. from the University of Oregon (1935), followed by an M.A. from

Indiana University (1936), and a Ph.D. in psychology from Stanford University (1938), under the direction of E. R. Hilgard. His early research started in the domain of individual differences; his first publication (Buxton & Humphreys, 1935) was titled “The Effect of Practice upon Intercorrelations of Motor Skills.” For his doctoral and some subsequent research, however, Humphreys made substantial contributions to the basic experimental learning literature regarding the nature of partial reinforcement (e.g., Humphreys, 1939).

Although Humphreys briefly held faculty appointments at Northwestern University, the University of Washington, and Stanford University, a substantial part of his research career has been devoted to military testing—from 1942 to 1945, as a member of the Army Air Forces Aviation Psychology Research Program during World War II (e.g., Humphreys, 1947), and from 1951 to 1957, as research director, Personnel Laboratory, Air Force Personnel & Training Research Center. In 1957, he joined the faculty at the University of Illinois at Urbana-Champaign as a professor of psychology. He remained in this position until his retirement in 1984. He continued to write on intelligence and related topics in his position as emeritus professor of psychology and education at the University of Illinois. Humphreys was the president of the Psychometric Society, president of two divisions (5 and 19) of the American Psychological Association, and served as editor of two journals: *Psychological Bulletin* and *American Journal of Psychology*.

INTELLIGENCE THEORY

The initial formulation of Humphreys’s theory of intelligence is contained in his 1961 Presidential Address to Division 5 (Evaluation and Measurement) of the American Psychological Association (Humphreys, 1962). In this address, Humphreys outlined a facet-based approach to human abilities that leads to a hierarchical structure of intelligence, with a broad, “general factor,” and less broad subsidiary factors. Although the ultimate conceptualization of a hierarchy of factors does not differ fundamentally from other such hierarchical theories, the foundation for Humphreys’s hierarchy is different—it incorporates the construct of test homogeneity with respect to facet-based analysis. That is, Humphreys demonstrated that

test construction, factor analysis, and ability theory are intertwined—ignoring one part of the equation jeopardizes the veracity of conclusions that can be made about the other parts of the equation. One primary implication of Humphreys's facet-based analysis was that the distinctions among intelligence, aptitude, and achievement tests are inherently problematic, partly because of the arbitrary dependence on test homogeneity for such classifications (e.g., Humphreys, 1974).

More recent formulations of Humphreys's theory of intelligence have focused on the nature and importance of the factor of general intelligence (e.g., Humphreys, 1979, 1985). His conceptualization of general intelligence owes much to early association theorists, such as G. Thomson, E. L. Thorndike, R. Tryon; but also to G. A. Ferguson's theory of learning and intelligence. The concept has been stated as follows: "Intelligence is the resultant of the processes of acquiring, storing in memory, retrieving, combining, comparing, and using in new contexts information and conceptual skills" (Humphreys, 1979, p. 115).

INTELLIGENCE AND RESEARCH METHODOLOGY

Humphreys's contributions to research methods are many and varied. Several of these contributions have proven to be integral to the interpretation of data from intelligence and ability investigations. Few modern intelligence theorists have been able to avoid the scrutiny of Humphreys's incisive methodological criticisms and reanalyses. Over the years, Humphreys has countered the methods and conclusions of, for example, R. B. CATTELL ("Critique of Cattell's 'Theory of fluid and crystallized intelligence: A critical experiment'"; Humphreys, 1967); A. R. JENSEN ("Jensen's theory of intelligence"; Humphreys & Dachler, 1969); J. R. Kirby and J. P. Das ("Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis"; Humphreys, 1978); and R. J. Sternberg ("Inadequate data in, attractive theory out; Humphreys, 1982). Humphreys also demonstrated that so-called Piagetian tests of intelligence share substantial variance with traditional intelligence quotient (IQ) tests (Humphreys & Parsons, 1979).

The Simplex. One of the core areas for Humphreys's attention has been the investigation of the

"simplex" structure of some types of correlational data. (Although the simplex pattern has a quantitative specification, the critical concept is that for any variable measured on multiple occasions [e.g., IQ], correlations between closely spaced testing periods are higher than for testing periods more distantly spaced. As a general rule, as the number of intervening testing periods increases, the correlations between initial and final level decrease.) As Humphreys noted in his presidential address to the Psychometric Society in 1959 (Humphreys, 1960), the simplex is a fundamental property of many types of longitudinal data—not just IQ measures, but also learning data and nonpsychological phenomena, such as the weather and height. A fundamental finding by Humphreys was that it is psychologically indefensible to subject simplex-patterned data to some methods of common factor analysis. Such analyses yield misleading findings of so-called early and late factors.

In the ensuing decades, Humphreys and his students devoted substantial study to the ubiquitous simplex-like patterning of multi-occasion correlation matrices. The research occurred in the context of intellectual development (e.g., Humphreys, Davy, & Park, 1985; Humphreys, Park, & Parsons, 1979) and the context of predicting academic grades during college (Humphreys, 1968).

Factor Analysis. In addition to the simplex, Humphreys and his colleagues have contributed to the methods for factor analysis, a backbone for investigation of intellectual abilities. Most notable, perhaps, is the development of a parallel analysis procedure (in collaboration with R. Montanelli), which allows for an objective, statistically based criterion for determining the number of factors underlying a matrix of correlations (Humphreys & Montanelli, 1975; Montanelli & Humphreys, 1976).

INTELLIGENCE AND PUBLIC POLICY

Never one to shy away from discourse on volatile issues, Humphreys was at the vanguard of psychologists willing to discuss implications of intelligence research on sex, race, ethnic group, and socioeconomic status differences. Humphreys contributed to the discussion of the heritability of intelligence (e.g., Humphreys, 1971), as well as to the definition and

assessment of bias in ability testing (Cleary et al., 1975). He documented a wide array of differences in academic achievement of different ethnic and racial groups, and presented a series of recommendations for remediation (Humphreys, 1988). Other issues related to public policy and intelligence were also discussed (e.g., Humphreys, 1989). While so many contemporary psychologists were content to offer only heated discussion, Humphreys's critical evaluation of group differences in intellectual abilities, similar to his evaluation of research methods in other domains, provided truly thoughtful and objective discussion of the issues.

BIBLIOGRAPHY

- BUXTON, C. E., & HUMPHREYS, L. G. (1935). The effect of practice upon intercorrelations of motor skills. *Science*, *81*, 441–442.
- CLEARY, T. A., HUMPHREYS, L. G., KENDRICK, S. A., & WESMAN, A. (1975). Educational uses of tests with disadvantaged students. *American Psychologist*, *30*, 15–41.
- HUMPHREYS, L. G. (1939). The effect of random alternation of reinforcement on the acquisition and extinction of conditioned eyelid reactions. *Journal of Experimental Psychology*, *25*, 141–158.
- HUMPHREYS, L. G. (1947). The program of printed test development. In J. P. Guilford & J. I. Lacey, (Eds.), *Army Air Forces Aviation Psychology Program Research Reports: Printed Classification Tests. Report No. 5* (pp. 13–24). Washington, DC: U.S. Government Printing Office.
- HUMPHREYS, L. G. (1960). Investigations of the simplex. *Psychometrika*, *25*, 313–323.
- HUMPHREYS, L. G. (1962). The organization of human abilities. *American Psychologist*, *17*, 475–483.
- HUMPHREYS, L. G. (1967). Critique of Cattell's "Theory of fluid and crystallized intelligence: A critical experiment." *Journal of Educational Psychology*, *58*, 129–136.
- HUMPHREYS, L. G. (1968). The fleeting nature of the prediction of college academic success. *Journal of Educational Psychology*, *59*, 375–380.
- HUMPHREYS, L. G. (1971). Theory of intelligence. In R. Cancro (Ed.), *Intelligence: Genetic and environmental influences* (pp. 31–42). New York: Grune and Stratton.
- HUMPHREYS, L. G. (1974). The misleading distinction between aptitude and achievement tests. In D. R. Green (Ed.), *The Aptitude-Achievement Distinction: Proceedings of the Second CTB/McGraw-Hill Conference on Issues in Educational Measurement* (pp. 262–285). Carmel, CA: CTB/McGraw-Hill.
- HUMPHREYS, L. G. (1978). Doing research the hard way: Substituting analysis of variance for a problem in correlational analysis. *Journal of Educational Psychology*, *70*, 873–876.
- HUMPHREYS, L. G. (1979). The construct of general intelligence. *Intelligence*, *3*, 105–120.
- HUMPHREYS, L. G. (1982). Inadequate data in, attractive theory out. *Journal of Educational Psychology*, *74*, 424–426.
- HUMPHREYS, L. G. (1985). General intelligence: An integration of factor, test, and simplex theory. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements, and applications* (pp. 201–224). New York: Wiley.
- HUMPHREYS, L. G. (1988). Trends in levels of academic achievement of blacks and other minorities. *Intelligence*, *12*, 231–260.
- HUMPHREYS, L. G. (1989). Intelligence: Three kinds of instability and their consequences for policy. In R. L. Linn (Ed.), *Intelligence: Measurement, theory, and public policy* (pp. 193–216). Urbana: University of Illinois Press.
- HUMPHREYS, L. G., & DACHLER, H. P. (1969). Jensen's theory of intelligence. *Journal of Educational Psychology*, *60*, 419–426.
- HUMPHREYS, L. G., DAVY, T. C., & PARK, R. K. (1985). Longitudinal correlation analysis of standing height and intelligence. *Child Development*, *56*, 1465–1478.
- HUMPHREYS, L. G., & MONTANELLI, R. G., JR. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, *10*, 193–206.
- HUMPHREYS, L. G., PARK, R. D., & PARSONS, C. K. (1979). Application of a simplex process model to six years of cognitive development in four demographic groups. *Applied Psychological Measurement*, *3*, 51–64.
- HUMPHREYS, L. G., & PARSONS, C. K. (1979). Piagetian tasks measure intelligence, and intelligence tests assess cognitive development: A reanalysis. *Intelligence*, *3*, 369–382.
- LINN, R. L. (ED.). (1989). *Intelligence: measurement, theory, and public policy*. Urbana: University of Illinois Press.
- MONTANELLI, R. G., JR., & HUMPHREYS, L. G. (1976). Latent roots of random data correlation matrices with squared multiple correlations on the diagonal: A Monte Carlo study. *Psychometrika*, *41*, 341–348.

PHILLIP L. ACKERMAN

HUNT, JOSEPH MCVICKER (1906–1991)

The American psychologist Joseph McVicker Hunt was born March 19, 1906, and raised on a farm in Scottsbluff, Nebraska, the son of R. Sanford and Carrie Pearl (McVicker) Hunt. He attended local public schools and the University of Nebraska in Lincoln, graduating in 1929. He remained there to study with J. P. GUILFORD, completing his M.A. in 1930, but then moved to Cornell University to work with Madison Bentley, finishing his Ph.D. in 1933. Hunt's early interests centered on personality, psychopathology, and the neurological as well as experiential underpinnings of personality disorders. To expand his training in this area, he did postdoctoral work at the New York Psychiatric Institute and at Worcester State Hospital in Massachusetts. Although Hunt's interest in personality functioning continued throughout his life, his best-known publication on this topic is the two volumes of *Personality and Behavior Disorders*, which he edited in 1944, and which became a standard reference text in the emerging field of clinical psychology for many years. In recognition of his achievements in this area, Hunt received in 1976 the Distinguished Contribution Award from the Division of Clinical Psychology of the American Psychological Association.

Hunt's first regular academic position was at Brown University, where he started research with rats on the effects of early rearing environments in order to examine psychoanalytic insights into the development of personality characteristics. During this time, the Yale group around Clark Hull became an important reference group for him, and he oriented his thinking toward the predominant behavioristic psychology.

Hunt's interest in intelligence grew out of his concern with the effects of early experience. Having recognized the extremely important influence of early experience on personality development, he also began to question the prevailing view of predetermined intellectual ability. After moving to the University of Illinois in 1951 and starting to examine the effects of child-rearing practices from a broad historical perspective, Hunt became impressed with the evidence for plasticity in intellectual development. This led him to question the assumptions that intelligence, particularly as measured by IQ tests, is genetically predetermined, impervious to the environmental circumstances during development, and unchanging in

character. These efforts resulted in the publication of his most influential book, *Intelligence and Experience* (1961).

This book crystallized a major shift in thinking about the nature of intelligence and its determinants. Hunt integrated a great deal of varied evidence to support a conception of intelligence in terms of *central processes* that undergo qualitative transformations during development and are quite open to the effects of experience. In this work, he brought together the theorizing of D. O. HEBB and of Jean PIAGET to present a view of intelligence as an information-processing system, subject to environmental influences during development at the level of neuroanatomical and neurological structuring as well as at the psychological level of motivation and attainment. Hunt suggested that a discrepancy between expectation and environmental demands serves as the major motivating force for intellectual effort, describing it as motivation inherent in information processing and action, or intrinsic motivation. He raised the possibility that the limit of human intellectual potential might be much greater than is typically attained, because the conditions of optimal discrepancy at different levels of development are not known.

This book became a focus for efforts to improve the intellectual functioning and educational achievements of children from poverty and general under-privilege and contributed to the vision embodied in Project Head Start. Hunt became an important figure in attempts to expand opportunities for early education of such children. He chaired the Presidential Task Force on Child Development, which produced the report "A Bill of Rights for Children" in 1967 and recommended Follow Through classrooms as well as neighborhood Parent and Child Centers in order to mobilize the broader support needed to foster children's educational achievement. Many of Hunt's writings dealing with early education and intellectual functioning were gathered into *The Challenge of Incompetence and Poverty* (1969).

Hunt believed in the relevance of research evidence for understanding human functioning, so he started an ambitious research program to determine the ingredients of stimulating and development-fostering environments during infancy. He found collaborators in Greece and Iran to help him evaluate the intellectual

development of infants being reared in orphanages where only minimal attention was being given to the infants' psychological needs. In order to examine the effects of introducing different kinds of environmental stimulation on development, Hunt thought that a new kind of assessment instrument was needed. The resulting scales (constructed in collaboration with Ina Č. Užgiris, 1975) were grounded in the Piagetian conception of qualitative transformation and differentiation of intelligence with development. They differ from standard intelligence tests in that individual performance is compared to sequential developmental landmarks rather than to norms achieved by chronological age groups.

On the basis of subsequent research, Hunt became convinced that a large part of the relevant environmental effects on intelligence are mediated by social facilitation of interest, curiosity, and trust in one's competence. Thus, indirectly, he was brought full circle to his original interest in personality characteristics, but now his interest was about their relevance to intellectual functioning. Hunt received in 1979 the Gold Medal of the American Psychological Foundation for his lifelong achievements.

Hunt's major contribution to the understanding of intelligence was his freeing of psychology from the assumptions of fixed and predetermined intellectual levels, largely impervious to postnatal experience. Without moving purely to a mentalistic plane, Hunt offered an alternative conception of intelligence as an information-processing system, grounded in the development of neural structures, responsive to stimulation available during the early years of development, and undergoing qualitative transformations in the years of growth in childhood. In the case of humans, he saw that the environments are very much self-selected or self-created, and, therefore, posited the importance of intrinsic motivation for intellectual development. This led him to urge research for optimal educational programs that would foster intrinsic motivation, self-confidence, curiosity, venturesomeness, a broadening of experience and the construction of intellectual competencies to deal with a wider range of problems and a general increase in problem-solving ability.

(See also: NATURE, NURTURE, AND DEVELOPMENT.)

BIBLIOGRAPHY

- HAYWOOD, H. C. (1992). Joseph McVicker Hunt (1906–1991). *American Psychologist*, 47(8), 1050–1051. Obituary.
- HUNT, J. MCV. (1944). *Personality and the behavior disorders* (2 vols.). New York: Ronald Press.
- HUNT, J. MCV. (1961). *Intelligence and experience*. New York: Ronald Press.
- HUNT, J. MCV. (1965). Intrinsic motivation and its role in psychological development. In D. Levine (Ed.), *Nebraska symposium on motivation* (Vol. 13, pp. 189–282). Lincoln: University of Nebraska Press.
- HUNT, J. MCV. (1969). *The challenge of incompetence and poverty: Papers on the role of early education*. Urbana: University of Illinois Press.
- HUNT, J. MCV. (1971). Intrinsic motivation: Information and circumstance. In H. M. Schroder & P. Suedfeld (Eds.), *Personality theory and information processing* (pp. 85–130). New York: Ronald Press.
- HUNT, J. MCV. (1971). Intrinsic motivation and psychological development. In H. M. Schroder & P. Suedfeld (Eds.), *Personality theory and information processing* (pp. 131–177). New York: Ronald Press.
- HUNT, J. MCV. (1974). A professional odyssey. In T. S. Kra-wiec (Ed.), *The psychologists* (Vol. 2, pp. 135–202). New York: Oxford University Press.
- HUNT, J. MCV. (1979). *Early psychological development and experience*. Worcester, MA: Clark University Press.
- UŽGIRIS, I. Č. (1979). Hunt, J. McVicker. In D. L. Sills (Ed.), *International encyclopedia of the social sciences* (Vol. 18, pp. 330–333). New York: Free Press.
- UŽGIRIS, I. Č., & HUNT, J. MCV. (1975). *Assessment in infancy*. Urbana: University of Illinois Press.

INA Č. UŽGIRIS

HYPERACTIVITY The term *hyperactivity* is commonly used to describe the behavior of troubled children. Despite the popularity of this label, a lot of confusion, conflict, and misunderstanding exists about it. Teachers, doctors, psychologists, and parents often use the term, but much controversy about its exact definition, measurement, and cause persists. Many questions remain, although some have been answered. Does it mean something is wrong with the brain? Is it the parents' fault? What can be done to treat it? Should it really be treated? Won't eating the right

foods make it go away? Isn't it usually outgrown anyway?

Hyperactivity has been used to mean several things—in the recent history of psychiatry and psychology, and this has been the cause of a good deal of the confusion. When used to describe a symptom, the term refers to a quantity and speed of physical movement that is substantially greater than that of most people of the patient's age. Any number of problems, illnesses, and disorders can cause the symptom of hyperactivity. The term can also be used as the name of a specific psychiatric or psychological disorder, which is currently known as ATTENTION DEFICIT HYPERACTIVITY DISORDER (ADHD). Patients displaying this disorder are characterized by a combination of symptoms that include problems with inattention, impulsivity, and hyperactivity. People who display the symptom of hyperactivity alone do not necessarily have the disorder itself, and people with this disorder do not necessarily display the symptom of hyperactivity.

HYPERACTIVITY AS SYMPTOM

There is no universally accepted way to define precisely how much activity constitutes hyperactivity. The diagnostic criteria for this symptom in the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders (DSM)* merely state that the behavior is "considerably more frequent than that of most people of the same mental age." Thus, the behavior must be assessed in comparison with others and cannot be evaluated concretely without a good bit of experience. What might appear to be excessive overactivity to a young, first-time parent may be regarded as normal to parents of several children, who have learned to expect a range of activity levels from their children. Hyperactivity as a symptom tends to vary in appearance, depending on a number of factors, including the situation the child is in, the age of the child, the amount of stimulation or distraction in the environment, and the child's familiarity with the setting.

To evaluate properly the possibility that a child is displaying the symptom of hyperactivity, both age and situational factors need to be taken into consideration. Activity levels that are normal and expected for an average 3-year-old would be considered very symptomatic if observed in a 10-year-old. The hyperactivity

symptom also manifests itself differently in respect to age. For those children who display hyperactivity as early as infancy, a history of irritability, colic, and sleeplessness is reported. Parents frequently report that during the toddler stage their child never really learned to walk but seemed to learn to run first, as if wound up and driven by a motor. They describe a "whirling dervish" sort of child who was constantly moving, getting into things, climbing, and jumping. Such behavior puts these children at considerably greater risk of accident and injury, with trips to doctors and emergency rooms something of a common occurrence.

As the child enters preschool and early elementary school, the symptom of hyperactivity becomes more apparent as teachers, well used to the normal range of childhood activity, note that these children have considerably more trouble settling down to quiet and sedentary activities than their classmates. Such children cannot seem to stay in their seats, often talk out of turn, and are easily frustrated. Normal attempts to help them to regulate better their activity levels meet with little success, resulting in frustration for adults and children alike. At this age and onward into early school years, movements that appear to have no good purpose are often observed as the child constantly fidgets, squirms, twists, taps, and wriggles hands, feet, and fingers. This quality of fidgetiness often continues into adolescence and beyond, though more noticeable overactivity and "hyperness" are often outgrown.

Situational factors also appear to have an influence on the hyperactivity symptom. In fact, some researchers believe that it is not the amount of overactivity that is the problem but, rather, that the overactivity occurs in inappropriate settings or situations (Copeland & Walraich, 1987). Some children appear appropriate and normal in some settings and hyperactive in others. For example, on the playground, a child with hyperactivity is not much different from others, since high amounts of activity are quite normal there. Sometimes differences between home and school behavior are observed. As a rule, these behaviors are most obvious in situations that call for quiet, seated, focused attention. They are more likely to emerge when the individual is bored or uninterested in what is going on and least likely to appear when the child is absorbed, interested, and stimulated to attend. Not surprisingly,

classroom, library, and lecture type settings seem to elicit these behaviors. In new or novel situations, such as initial visits to a new doctor or first meetings with a new teacher, hyperactive behaviors are often absent or minimal. Children who can sit for hours mastering a new, exciting video game may be unable to focus on a schoolroom activity for more than a few minutes.

To assess and measure the amount of hyperactivity an individual displays, clinicians have developed a number of measurement tools. Such tools are an important means of turning observations of behavior based on impressions into more objective, measurable descriptions. Just as a thermometer measures the temperature of a room and turns that into a number, so these tools help to quantify activity levels and other behavior. Two basic types of tools used to measure this behavior are rating scales and movement-monitoring devices.

Rating scales have been developed for teachers and parents to use when observing a child's behavior. A number of characteristics, such as restlessness or failure to finish tasks, are rated on a zero (none) to four (very much) basis, and the total score is the sum of each of the ratings on each characteristic. With proper treatment, these total scores tend to drop rapidly, indicating a reduction in the symptom.

Electromechanical movement-monitoring devices, which actually measure the amount of physical movement a person makes, are sometimes used to study the problem. Usually attached to a child's wrist or ankle, these watchlike electronic devices record the number of movements made during a set period of time. Devices such as these revealed that hyperactive children are no more active than other children on the playground but much more active than others in the classroom. These tools are important for several purposes, including measuring the effect possible treatments for hyperactivity and comparing the level of hyperactivity children may display in different settings.

CAUSES OF HYPERACTIVITY

Once it has been determined that someone is displaying the hyperactivity symptom, the next step toward treatment is to seek the symptom's root cause. Any number of medical, neurological, and psychological disorders can cause hyperactivity. The list of prob-

lems that can cause this symptom is a relatively long one, though many of the potential causes are quite rare. It is important to consider other potential causes when hyperactivity is observed so that a treatment that addresses the root cause can be provided. Several types of psychiatric, psychological, neurological, and environmental problems can result in, or be associated with, hyperactivity. They include depression, anxiety disorders, learning disabilities, mental retardation, autistic disorder, chaotic social environments, Tourette's syndrome, seizure disorders, lead poisoning, and certain medications. Sorting through these root causes is not a simple matter and requires a good deal of training and experience. Many doctors and school counselors routinely consult with specialists in psychiatry, neurology, clinical psychology, and special education to determine what lies behind the symptom and decide on the correct treatment.

Several authors have proposed that dietary factors, including food allergies, sugar hypersensitivity, and food additives, are a common root cause of hyperactivity. Very elaborate and often expensive nutrition systems that seek to eliminate these food components have been developed to treat the problem, and whole families have emptied out their cupboards and altered their nutritional habits in an effort to help a hyperactive family member. While a certain number of families who have tried these diets have reported them to be helpful, attempts to determine scientifically if such approaches really work have not been promising.

To test whether or not a new approach is actually helpful, researchers must carefully construct a series of experiments. In this process, an essential step is to compare the outcome of people receiving the treatment with a group receiving a seemingly very similar type of treatment, called a placebo treatment. A placebo is medicine or treatment that is inert, or without power, one that lacks the crucial active agent but otherwise is very similar to the "real" treatment. It is not uncommon for 10–25 percent of patients to improve in response to placebos. This response is thought to be due to the psychological effects of wanting to improve and the "power of suggestion." It is well documented that people may respond to all sorts of ineffective treatments as long as they believe that the treatment has the power to cure them. To prove that a particular treatment is actually effective, rates of improvement

from the actual treatment must be significantly greater than those observed as a result of placebo treatment alone. Research of this nature has failed to demonstrate that these dietary approaches are any more effective than placebo treatment. That is, about as many children improve when put on a placebo diet as when put on a diet that limits food additives, refined sugar, or food colorings. While a certain few children may in fact be sensitive to such substances, this is clearly a rare root cause of hyperactivity (Goldstein & Goldstein, 1990).

HYPERACTIVITY AS DISORDER

Many of the children and adolescents who display the symptom of hyperactivity also have significant problems with inattention and impulsivity. They tend to be easily distracted, seem not to listen to others, and often appear to act without thinking through their actions. They have difficulty completing schoolwork in an organized way, frequently move quickly from one thing to another, finishing nothing. This disorder has been called many things, including "minimal brain syndrome," "minimal brain dysfunction," "hyperkinesis," "hyperkinetic syndrome of childhood," "attention-deficit disorder," and "hyperactive syndrome." The current *DSM* term for this disorder is, as noted, attention deficit hyperactivity disorder.

Progress in medicine, psychiatry, and psychology often comes as a result of identifying subtypes within a diagnostic group, that is, a group whose members uniquely share certain symptoms and lack others. This subtyping allows the development of specialized treatments that are more effective for one subtype of a disorder than another. A number of attempts to identify unique subtypes within ADHD in the past have failed to yield such diagnostic specificity. Modern research has been focused on the importance of the hyperactivity symptom in ADHD. In one study, ADHD children with hyperactivity (ADHD + H) had considerably more problems with self-control, impulsivity, and aggression than did ADHD children who lacked prominent hyperactivity (ADHD - H). ADHD + H children also had more problems getting along with their peers, were more often suspended from school, and tended more toward delinquent and antisocial activities than did those without hyperactivity. The

ADHD - H children were described as equally inattentive but significantly more confused, daydreamy, listless, apathetic, and lost in thought than their ADHD + H peers (Barkley, DuPaul, & McMurray, 1990). Earlier studies had found that ADHD + H children have significantly more trouble getting along with other children than ADHD - H youngsters, with more ADHD + H voted "least liked" child in the classroom (Carlson et al., 1987). The tendency in ADHD + H children to interrupt other children's games and conversations, their failure to wait their turn during play, their problems responding to social cues, and their tendency to be overly loud and aggressive lead to such peer rejection.

While evidence continues to accumulate that ADHD + H and ADHD - H may represent separate but overlapping disorders, there is not yet much evidence that this has strong implications for treatment. Some researchers have suggested that ADHD + H children are more likely to respond to stimulant medications than ADHD - H children, but firm conclusions in this area can not yet be drawn (Cantwell & Baker, 1992).

INTELLIGENCE AND OTHER CORRELATES OF HYPERACTIVITY

Whether hyperactivity occurs as an isolated symptom or as part of a more pervasive ADHD disorder, this problem has profound effects on the child and the family. For any given child, it is impossible to say to what degree personality characteristics and behavior are the result of inborn temperament or of experiences in life. At the same time, it is impossible to say just how much of the family conflict and academic failure can be prevented or avoided if the problem is appropriately treated. Nonetheless, it is clear that these symptoms and resulting difficulties are related and have a clear impact on the family, the child's self-esteem, and ultimately on society.

Ordinarily, by the time a hyperactive child reaches the age of 6 or 7, he or she will have repeatedly experienced a good deal of exasperation, irritation, and rejection from parents, teachers, siblings, playmates, and others. The world is seen by such a child as a negative, punishing, rejecting place, and this worldview has a tremendous effect on the child's sense of

self-worth and opinion of others. Such children come to expect criticism and feel helpless to please others. A rejected child is more likely to feel frustrated and to act aggressively. Thus, a vicious cycle of failure, frustration, rejection, and anger is formed. Early schooling requires a great deal of boring repetition, practice, and patience. A child who consistently fails to succeed in such an environment because of inattention, impulsivity, or hyperactivity falls further and further behind and has terrible difficulties mastering the basics of reading, spelling, and arithmetic. Many of these children regard themselves as stupid and are treated as a "retard" by their classmates, despite their normal IQs. They understandably see others as mean, demanding, and unhelpful. Reaching junior high only amplifies these problems, as the adolescent moves from classroom to classroom, from teacher to teacher, and begins to get more homework requiring planning, self-direction, and organization. Even if hyperactivity diminishes as the child matures, it is very difficult, if not impossible, to catch up academically and to undo the effects on self-esteem and relationships with others. Many such children simply give up and, totally soured by the school experience, drop out.

Similar effects are seen in the family. The hyperactive child's problems add greatly to the tensions of normal family life. These children seem impossible to control and do not seem to listen or be affected by attempts to discipline. Such children are disappointing to their parents, who feel inept, angry, guilty, and helpless. Attempts to control their behavior through discipline may become harsh, inconsistent, and extreme. Other children in the family may be viewed as "good" and the hyperactive child as "bad," contributing further to tensions, resentment, and conflicts within the family and to the child's sense of worthlessness. By the time adolescence is reached, even if the child's hyperactivity diminishes, the foundation of conflict within the family has been established, and tends to worsen as the normal rebelliousness of the teenage years emerge. Many such young people simply give up and run away.

It was long thought that these problems were usually outgrown as part of the normal process of maturation, but a great deal of evidence now shows problems extending into the late teen years and beyond. For example, R. G. Klein and S. Mannuzza

(1991) reviewed the long-term outcome of hyperactivity in childhood. Their review included studies of patients diagnosed in childhood as hyperactive who were reassessed in adolescence and adulthood. It revealed that a substantial subgroup who remain hyperactive develop antisocial personality disorders and serious drug-abuse patterns, both of which are related closely to the emergence of criminal behavior, arrests, and incarceration. At present, no direct means of identifying which subgroup of hyperactive children will become antisocial adults is available. Many clinicians believe that treating hyperactivity in a multimodal fashion, with a combination of medications, psychotherapy, family therapy, and academic assistance, can prevent this spiral into antisocial behavior, but research supporting this belief is still sparse. It is clear that such treatment improves concentration, decreases impulsivity and hyperactivity, improves peer relationships, and thus holds the hope of reducing negative effects of hyperactivity on school and family life. Whether or not such treatment impacts the long-term outcome remains unclear, and research along these lines is ongoing.

THE OUTLOOK

Hyperactivity and related problems seriously affect 6–9 percent of children at some point in their lives. Research investigating its causes, typical course, measurement, and treatment continues. Despite hundreds of scientific studies, there is still a great deal we do not know about this surprisingly common malady. New and highly sophisticated technical methods of investigation in neurology, including techniques for scanning and monitoring the brain's electrical and chemical functioning, will likely allow further understanding of the nature of the symptom itself and may lead to even better treatment strategies. Improved methods for formally assessing the mental and psychological causes and effects of hyperactivity are also being developed by clinical neuropsychologists and educators. Incomplete but growing evidence suggests that hyperactivity may well be a significant root cause of many forms of alcohol and substance abuse, delinquency, educational failure, child abuse, spousal abuse, depression, violence, criminality, and personality disorders. Future investigation of the relationships be-

tween hyperactivity and each of these serious social problems will surely shed new light on many of these issues, with the exciting prospect that some of them can be prevented from emerging with proper treatment. Scientific understanding of hyperactivity is increasing rapidly and with it grows the hope that its many negative effects can be avoided.

(See also: LEARNING DISABILITY.)

BIBLIOGRAPHY

- AMERICAN PSYCHIATRIC ASSOCIATION. (1987). *Diagnostic and statistical manual of mental disorders* (3rd, rev. ed.). Washington, DC: Author.
- BARKLEY, R. A., DUPAUL, G. J., & McMURRAY, M. B. (1990). Comprehensive evaluation of attention deficit disorder with and without hyperactivity as defined by research criteria. *Journal of Consulting and Clinical Psychology, 58*, 775-789.
- CANTWELL, D. P., & BAKER, L. (1992). Attention deficit disorder with and without hyperactivity: A review and comparison of matched groups. *Journal of the American Academy of Child and Adolescent Psychiatry, 31*, 432-438.
- CARLSON, C. L., LAHEY, B. B., FRAME, C. L., WALKER, J., & HYND, G. W. (1987). Sociometric status of clinic-referred children with attention deficit disorder with and without hyperactivity. *Journal of Abnormal Child Psychology, 15*, 537-547.
- COPELAND, L., & WOLRAICH, M. L. (1987). Attention deficit hyperactivity disorder. In M. L. Wolraich, (Ed.), *The Practical Assessment and Management of Children with Disorders of Development and Learning*. Chicago: Year Book Medical.
- GOLDSTEIN, S., & GOLDSTEIN, M. (1990). *Managing attention disorders in children: A guide for practitioners*. New York: Wiley.
- KLEIN, R. G., & MANNUZZA, S. (1991). Long-term outcome of hyperactive children: A review. *Journal of the American Academy of Child and Adolescent Psychiatry, 30*, 383-387.

MARK F. WARD

I

IDIOTS SAVANTS See SAVANTS.

ILLITERACY Illiteracy refers to the inability to read and write in order to function in society. Estimates of the number of Americans who are illiterate vary widely, from 27 million to 72 million people, depending on the standard one uses to define the ability to read and write well enough to function in society. A number of factors are related to the incidence of illiteracy, including amount of education, extent of language knowledge, methods used to teach reading and writing, level of intelligence, and presence of learning disabilities. Treatments designed to combat illiteracy vary as a function of program goals and clientele served. For individuals who are out of school, approaches include volunteer and competency-based programs, workplace programs, and family programs. For those in school, approaches include remedial and preventive programs.

HOW MANY AMERICANS ARE ILLITERATE?

Over the years, the level of reading and writing ability deemed necessary to function in American society, along with the ways in which those levels were measured, changed (see Cook, 1977; Sticht, 1988).

There have also been considerable changes in the number of adults classified as functionally illiterate.

During the 1940s, the Census Bureau began to collect data on the number of years of schooling completed. Having completed less than five years of education became the standard used to define functional illiteracy. Based on that standard, nearly 14 percent (10 million) of adults over 25 years of age were functionally illiterate in 1940. By 1960, about 8 percent (8 million people) were reported to be the same. By then, analysts had begun to argue that less than an eighth-grade level of education put individuals at risk for functioning in society. Using that standard, the number of functionally illiterate adults rose to 22 million. More recently, a high school level of education has been used as a measure of what is required for success in a technologically complex society. According to 1980 census figures, more than 50 million out-of-school American adults lacked a high school diploma.

Results from the National Assessment of Educational Progress (Kirsch & Jungeblut, 1986; NAEP, 1985) are yet another way to illustrate how standards affect estimates of illiteracy. According to these data, less than 1 percent of American adults aged 21 to 25 years of age were unable to sign their name. About 5 percent could not locate factual information in a short newspaper article. Approximately 25 percent were un-

able to write a letter explaining that a mistake has been made in a billing charge. About 40 percent could not answer a question based on a graph. More than 60 percent could not understand the main idea of an editorial from *The New York Times*. More than 90 percent were unable to describe the difference between two types of employee fringe benefits after having read about each.

ILLITERACY AS A CONTINUUM

The easiest way to think about the problem of illiteracy and those who experience it may be to think of a continuum, one based on the knowledge and skills that people have, along with the kinds of literacy tasks they are and are not able to do (Chall, 1987).

At one end of the continuum would be those who are unable to read or write at all, or those who can read only the simplest of signs, labels, and instructions—not enough for informed citizenship or most jobs. About 27 million adults would be considered illiterate according to this standard. Many people in this group have long histories of difficulties in school. Included in this group as well are individuals with limited facility with the English language, most of whom are also illiterate in their native language.

In the middle of the continuum are individuals who have not yet learned to use their reading and writing skills as tools for learning. This group reads somewhere between the fourth- and eighth-grade levels, and is able to deal with tasks such as applications, simple instructions, and short newspaper articles. Their knowledge of words and meanings is fairly limited, however, and they experience problems with tasks that require coordination of two or more pieces of information, usually because of lack of fluency in reading and writing. Using this as our standard, 18 million more American adults would be classified as illiterate.

Toward the other end of the continuum are individuals who need help because they lack the knowledge and skills necessary to compete in today's job market. Many of these individuals have completed high school, but still lack the linguistic and conceptual knowledge and skills needed to read and write texts at advanced levels (Hirsch, 1987). If we use this as our standard, 27 million more American adults would be classified as illiterate.

See Chall (1983) for additional information on the features of this continuum, and the reports from the National Assessment of Educational Progress (e.g., Kirsch & Jungeblut, 1986; NAEP, 1985) for additional information on proficiency levels.

CAUSES AND CORRELATES OF ILLITERACY

Why do so many Americans fail to acquire the reading and writing proficiency required for their needs? The most critical factor is education. Although some children and adults appear to learn to read and write without instruction, the vast majority acquire literacy skills because they have been taught and have had the necessary practice. The best evidence for this is the international study of fifteen countries directed by Robert L. Thorndike in 1973. Countries that provided more education had higher reading levels, and within each country, those who attended school longer had higher reading achievement. Number of years of schooling was found to be positively related to the literacy performances of the young adults in the National Assessment study as well (Kirsch & Jungeblut, 1986).

Insufficient knowledge of the English language is another factor underlying the reading and writing difficulties of American children and adults. Technically, difficulty in reading and writing a second language is not considered a problem of literacy but of second-language learning. However, many who need to learn spoken and written English have limited schooling and literacy skills in their first language. According to 1987 U.S. Department of Education figures, English as a second language (ESL) students made up about one-third of the population of adult basic education students.

Methods of teaching are also related to the incidence of illiteracy. For more than seventy years the research on methods of teaching beginning reading has found that direct and systematic instruction including letter-sound correspondences produces better results than a method that focuses mainly on the meaning of the text (Adams, 1990; Chall, 1967). Moreover, special help given early to those who have difficulty results in higher reading achievement (Pinnell, 1989; Slavin, 1991).

Another strong correlate of literacy development is verbal intelligence. In fact, the correlation of scores on a test of reading comprehension and a paper-and-pencil intelligence test is as high as that between two comprehension tests (Thorndike, 1973–1974). It is important to note, in this regard, that reading and verbal intelligence are not as highly related at the beginning stages of reading development as they are at later stages. Chall (1967, 1983) reports that the best predictor of early reading achievement (up to grade three) is facility with the phonemic and print aspects of reading. Beginning with the fourth grade, verbal intelligence—and, in particular, knowledge of word meanings—takes on greater importance. Thorndike (1973–1974) found the same. Once decoding is mastered, knowledge of word meanings is the best predictor of reading comprehension. This is so because materials written at the fourth-grade level and higher include more unfamiliar, abstract, and literary words. Also, children who are good beginning readers tend to read more, thereby spurring further growth in their vocabulary knowledge and comprehension (see Stanovich, 1986), and ultimately, their verbal IQ.

Extreme difficulty in learning to read and write occurs also among children and adults who have normal or higher intelligence, good language, and adequate schooling. An estimated 10 to 15 percent of the population have such difficulty (Carroll & Chall, 1975)—called, at different times, learning disability, specific language disability, reading disability, and dyslexia. In such cases, the problem is usually with the phonological and print aspects of reading, although these early difficulties often bring a delay in the more advanced aspects of reading and writing. Estimates of the number of adults at the lowest literacy levels who are learning disabled run as high as 75 percent (Keefe & Meyer, 1988).

WHAT IS BEING DONE?

Programs to deal with the problem of illiteracy are quite varied in terms of their approaches and goals and the clientele they are designed to serve.

Volunteer programs often employ one-on-one instructional techniques, and include efforts at the national level (like Laubach Literacy International and Literacy Volunteers of America) as well as at the local church

and community level. These programs tend to serve the adults with the most limited reading and writing skills, despite the fact that even the best funded among them is able to offer only about ten hours of training to their volunteers (Chall, Heron, & Hilferty, 1986). It has been estimated that only about 3 percent of the adults who need help end up being reached through these efforts (Mikulecky, 1990), and little research has been conducted on the effectiveness of the methods that these programs use.

Competency-based programs serve students at higher literacy levels than the volunteer programs, often with the goal of helping them obtain a GED or High School Equivalency Diploma. Instruction in competency-based programs can occur on an individual basis or in groups. The curriculum is usually divided into units, and students work only on those units on which they have yet to demonstrate mastery. Overall, these programs appear to reach about 50 percent of those who do not complete high school, although only about 40 percent of those who study for the GED receive it (Kirsch & Jungeblut, 1986).

Concern about illiterate workers has led many corporations to initiate *workplace literacy programs* (Sticht, 1988). A recent estimate is that employers provide training each year for 15 million adults (Mikulecky, 1990). Research shows that specific literacy skills can be developed successfully in particular job-related contexts (Sticht, 1988).

Family literacy programs have their basis in the strong positive relationship that exists between parents' educational levels and the achievement of their children (Coleman et al., 1966). Among the 300 or so programs that exist nationwide, all share the goal of improving the literacy of parents and children by bringing them together for learning activities. At present, little research has been conducted on the effectiveness of this approach. The Parent and Child Education (PACE) program in Kentucky does report promising results, however (Heberle, 1990).

Policymakers warn that more than one-third of American children have “the deck stacked against them” before they even enter school (Hodgkinson, 1991). As such, *school-based solutions* to the problem of illiteracy, although hardly new, remain a top priority. Research shows that today's disadvantaged child who reads below grade level and who fails a grade is likely

to be tomorrow's high school dropout (Slavin, 1991). When disadvantaged children are provided with good literacy instruction, however, they learn to read and write as well as their more advantaged peers (Chall, Jacobs, & Baldwin, 1990). Moreover, adolescents who lag far behind in reading and writing development can be helped to close that gap while they are still in school (Curtis, 1993).

A school-based reading improvement program at Boys Town that uses individualized diagnosis and group instruction (8 students to 1 teacher) has resulted in gains of one year for each semester of instruction. Courses in the program are tailored to students' diagnostic profiles, stressing word recognition and decoding, fluency, knowledge of word meanings, reading comprehension and study skills, and so forth, depending on students' strengths and needs (Father Flanagan's Boys' Home, 1993).

Similar gains of one year per semester (20 hours of instruction) have been found by the Harvard Adult Reading Center, where adult students are given individual diagnoses and instruction based on their reading patterns (Chall, 1991).

Practices that reviewers of research have found to be effective for promoting literacy improvement include attending a school with extensive remedial programs, taking an extra reading class during the school day, and attending summer school (Braddock & McPartland, 1993). Less effective practices include before- or after-school coaching, mentoring programs, and peer tutoring—perhaps because these kinds of programs are not focused directly on reading and writing needs.

CONCLUSIONS

Although arguments persist about the actual number of Americans who are illiterate, few would disagree with the conclusion that the status of literacy in this country is far from what it needs to be, particularly given the growing complexity of everyday life (Chall, 1987). National assessments have begun to provide reliable data on strengths and needs in reading and writing, for both children and adults. A number of different approaches for promoting literacy are being tried. The next step will need to be more research effort directed toward establishing which pro-

grams work for which groups, for what goals, and at what cost.

(See also: LANGUAGE AND INTELLIGENCE.)

BIBLIOGRAPHY

- ADAMS, M. J. (1990). *Beginning to read*. Cambridge, MA: MIT Press.
- BRADDOCK, J. H., & MCPARTLAND, J. M. (1993). Education of early adolescents. *Review of Research in Education*, 19, 135–170.
- CARROLL, J. B., & CHALL, J. S. (EDS.). (1975). *Toward a literate society*. New York: McGraw-Hill.
- CHALL, J. S. (1967). *Learning to read: The great debate*. New York: McGraw-Hill.
- CHALL, J. S. (1983). *Stages of reading development*. New York: McGraw-Hill.
- CHALL, J. S. (1987). Developing literacy . . . in children and adults. In D. Wagner (Ed.), *The future of literacy in a changing world*. London: Pergamon Press.
- CHALL, J. S. (1991, October). *Patterns of adult reading*. Paper presented at the Conference on Literacy and Adults with Learning Disabilities, New York.
- CHALL, J. S., HERON, E., & HILFERTY, A. (1987). Adult literacy: New and enduring problems. *Phi Delta Kappan*, 69, 190–196.
- CHALL, J. S., JACOBS, V. A., & BALDWIN, L. E. (1990). *The reading crisis: Why poor children fall behind*. Cambridge, MA: Harvard University Press.
- COLEMAN, J. S., CAMPBELL, E. A., HOBSON, C. J., MCPARTLAND, J., MOOD, A. M., WEINFELD, F. D., & YORK, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Government Printing Office.
- COOK, W. D. (1977). *Adult literacy education in the United States*. Newark, DE: International Reading Association.
- CURTIS, M. E. (1993). *Reading diagnosis and intervention for at-risk adolescents*. Boys Town, NE: Boys Town Reading Center.
- FATHER FLANAGAN'S BOYS' HOME (1993). *The Boys Town Reading Curriculum*. Boys Town, NE: Author.
- HEBERLE, J. (1990). PACE: Parent and child education in Kentucky. *Family Resource Coalition Report*, 3, 13.
- HIRSCH, E. D. (1987). *Cultural literacy*. Boston: Houghton-Mifflin.
- HODGKINSON, H. (1991). Reform versus reality. *Phi Delta Kappan*, 73, 9–16.

- KEEFE, D., & MEYER, V. (1988). Profiles of and instructional strategies for adult disabled readers. *Journal of Reading*, 31, 614–619.
- KIRSCH, I. S., & JUNGBLUT, A. (1986). *Literacy: Profiles of America's young adults*. Princeton, NJ: Educational Testing Service.
- MIKULECKY, L. (1990). National adult literacy and lifelong learning goals. *Phi Delta Kappan*, 72, 304–309.
- NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS (1985). *The reading report card*. Princeton, NJ: Educational Testing Service.
- PINNELL, G. S. (1989). Reading recovery: Helping at-risk children learn to read. *Elementary School Journal*, 90, 161–183.
- SLAVIN, R. E. (1991). Chapter 1: A vision for the next quarter century. *Phi Delta Kappan*, 72, 586–592.
- STANOVICH, K. E. (1986). Matthew effects in reading: Some consequences of individual differences in the acquisition of literacy. *Reading Research Quarterly*, 21, 360–407.
- STICHT, T. G. (1988). Adult literacy education. *Review of Research in Education*, 15, 59–96.
- THORNDIKE, R. L. (1973). *Reading comprehension education in fifteen countries*. International studies in Evaluation III. Stockholm: Almqvist & Wiksell.
- THORNDIKE, R. L. (1973–1974). Reading as reasoning. *Reading Research Quarterly*, 9, 135–147.

JEANNE S. CHALL
MARY E. CURTIS

IMAGERY Imagery refers to the mental depiction or re-creation of people, objects, and events that are not actually present. By closing one's eyes and generating the appropriate image, for example, one can imagine the face and voice of a best friend, how appliances are arranged in a kitchen, or the manner in which an elephant runs. The experience of imagining something is similar in many respects to that of actually perceiving something, and, in some cases, the image can be exceptionally clear and vivid.

Imagery can also be used to envision things that are merely hypothetical. For example, a person can imagine a creature consisting of the head of a chicken and the body of a lion, and even explore how that creature might behave—although nothing of the sort actually

exists. Imagery can thus be used not only to retrieve past experiences, as when recalling how one's room looked when one was a child, but also to explore new possibilities.

There are many practical uses of imagery in everyday life. One can imagine the best route for getting to work or school, novel ways of using tools, or creative things to do on a vacation. Imagery can even be helpful in such common activities as cooking a meal. For example, one might imagine the likely texture, taste, and appearance of a favorite chocolate cake if the recipe were modified to include three additional egg whites.

Clearly, imagery is of considerable value in a variety of activities. But how, exactly, does imagery contribute to what we think of as intelligence? What aspects of intelligence are influenced by imagery? To address these questions, we consider some of the most important functions of imagery, which have been explored in recent experimental studies.

FUNCTIONS OF IMAGERY

Mental Image Scanning. When one is asked about the physical characteristics of something, such as its size or features, it often helps to retrieve an image and to imagine scanning it. For example, suppose someone asked you whether Abraham Lincoln had bushy eyebrows. Although few people would have ever learned this as an explicit fact about Lincoln, most could answer the question by imagining that they were looking at a picture of Lincoln and inspecting his eyebrows.

Retrieving and scanning an image facilitates what is called *crystallized intelligence*—the ability to recall useful information when needed. How quickly this can be done depends on several factors. Studies have found that the time it takes to imagine scanning from one feature of an image to another increases in proportion to the distance scanned, as if one were actually scanning a real picture or object (Kosslyn, Ball, & Reiser, 1978). In an image, as in a real object, it takes longer to find a small feature than to find a large feature (Kosslyn, 1975). For example, it is easier to tell what an elephant's eyes look like if you imagine that the elephant is standing next to you, as opposed to standing on a distant hill. These and other studies on mental image scanning are reviewed in Kosslyn (1980).

Mental Rotation. Imagery can also be used to mentally rotate objects or patterns. This serves three functions. First, it can help one to identify pictures or objects that appear at unfamiliar orientations, as when a photograph is turned around. Second, it enables one to mentally align objects that are at different orientations, to see how their shapes or other features might correspond. Third, it can help one to become properly oriented within one's environment. For example, if you were lost, and remembered looking at a map that was oriented the wrong way, you could imagine turning the map around in order to use it. These imagery skills bear on what is called *fluid intelligence*—the ability to see relationships and solutions to problems that were not previously learned.

In general, mental rotations are performed at a constant rate, with larger angles of rotation requiring more time to complete (Shepard & Metzler, 1971). With practice, most people can become proficient at using mental rotation, irrespective of the complexity of the imagined pattern or form (Cooper & Podgorny, 1976). In addition to mental rotation, other types of mental transformations have been explored, including imagined changes in the size and shape of objects. Studies on imagined rotations and transformations are reviewed in Shepard and Cooper (1982).

Mental Extrapolation. Another important function of imagery is to anticipate continuations of motions. For example, if a moving object passes momentarily out of view, one can imagine where and when the object will reappear by imagining how its motion would continue. These *mental extrapolations* are especially useful in complex activities, such as driving a car or performing in an athletic event, where one needs to anticipate and remember the consequences of one's actions. They bear on an aspect of intelligence known as *motor skills*.

Recent studies have explored the mental extrapolation of movement and its effect on one's memory. When a stimulus display implies a simple rotation of a pattern, an observer's memory for the orientation of the pattern is shifted forward, in the direction of the implied rotation. This phenomenon, called *representational momentum* (Freyd & Finke, 1984), occurs with both auditory and visual stimuli. As with physical momentum, representational momentum is proportional

to the implied velocity of the motion. These memory shifts can also occur with static stimuli, such as photographs. Freyd (1987) has suggested that the images we form have an inherent dynamic quality, which we can use to anticipate future events.

Creative Mental Synthesis. One of the most creative uses of imagery is to imagine putting old things together in new ways, to see what might result. Often, the discoveries are quite surprising and unexpected. For example, suppose you had a basketball, a yardstick, a flat sheet of plywood, and some glue. What kinds of interesting things could you make by combining these objects? The ability to imagine new combinations of things is called *mental synthesis*, and it plays an important role in *creative intelligence*.

Studies on mental synthesis have shown that most people are able to make creative discoveries by using imagery. When instructed to imagine combining letters, numbers, and other simple patterns, people can often come up with strikingly creative patterns and symbols (Finke & Slayton, 1988). When imagining combinations of three-dimensional object parts, people can often discover new inventions that could potentially be developed and marketed (Finke, 1990). These images can also be interpreted in more abstract ways, representing, for example, novel concepts in scientific fields. Studies on mental synthesis and their creative implications are reviewed in Finke (1990).

Problem Solving. A further way in which imagery contributes to intelligence is through the use of images to solve various types of problems. In trying to solve a mystery, for instance, it helps to visualize how all the clues could be put together. This is an example of what is called *convergent thinking*, the ability to see the underlying structure that connects many things.

Another type of problem-solving skill that imagery can facilitate is that of finding unusual uses for common objects. For example, one might imagine using a brick as a paperweight, a foot warmer, or an emergency hammer. This ability to think of unusual or unconventional possibilities is called *divergent thinking*.

Imagery can also be useful in helping one to construct *mental models* of physical or conceptual systems (Johnson-Laird, 1983). Suppose, for example, one were trying to understand how a car engine worked. One could construct an image representing a possible

model for how the different parts of the engine might function and then evaluate the model. Mental models can help one understand difficult concepts, such as the theory of relativity, by incorporating visual or spatial analogies that are easier to comprehend. Reviews of the various functions of imagery in problem solving can be found in Finke, Ward, and Smith (1992).

CONCLUSIONS

What is it about imagery that makes it so different from other forms of mental representation and allows it to make these distinctive contributions to intelligence? First, most theories of imagery assume that images draw on perceptual knowledge and relations that are not normally available when one is merely retrieving facts about something (Finke, 1979). Second, images represent this information in a coherent, accessible manner that allows new and often complex relations to be recognized quickly and efficiently (Kosslyn, 1980). Third, images exhibit a flexibility that is often absent in other forms of mental representation (Finke, 1989). There is even evidence that certain parts of the brain are specifically designed to generate images and to discover the intricate relationships and possibilities they depict (Kosslyn, 1987).

BIBLIOGRAPHY

- COOPER, L. A., & PODGORNÝ, P. (1976). Mental transformations and visual comparison processes: Effects of complexity and similarity. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 503–514.
- FINKE, R. A. (1989). *Principles of mental imagery*. Cambridge, MA: MIT Press.
- FINKE, R. A. (1990). *Creative imagery: Discoveries and inventions in visualization*. Hillsdale, NJ: Erlbaum.
- FINKE, R. A., & SLAYTON, K. (1988). Explorations of creative visual synthesis in mental imagery. *Memory & Cognition*, 16, 252–257.
- FINKE, R. A., WARD, T. B., & SMITH, S. M. (1992). *Creative cognition: Theory, research, and applications*. Cambridge, MA: MIT Press.
- FREYD, J. J. (1987). Dynamic mental representations. *Psychological Review*, 94, 427–438.
- FREYD, J. J., & FINKE, R. A. (1984). Representational momentum. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 126–132.
- JOHNSON-LAIRD, P. N. (1983). *Mental models: Towards a cognitive science of language, inference, and consciousness*. Cambridge: Cambridge University Press.
- KOSSLYN, S. M. (1975). Information representation in visual images. *Cognitive Psychology*, 7, 341–370.
- KOSSLYN, S. M. (1980). *Image and mind*. Cambridge, MA: Harvard University Press.
- KOSSLYN, S. M. (1987). Seeing and imagining in the cerebral hemispheres: A computational approach. *Psychological Review*, 94, 148–175.
- KOSSLYN, S. M., BALL, T., & REISER, B. J. (1978). Visual images preserve metric spatial information: Evidence from studies of image scanning. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 47–60.
- SHEPARD, R. N., & COOPER, L. A. (1982). *Mental images and their transformations*. Cambridge, MA: MIT Press.
- SHEPARD, R. N., & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703.

RONALD A. FINKE
JENNIFER J. FREYD

IMAGINATION See INSIGHT; INTUITION.

INDIVIDUAL TESTS Individual tests of intelligence are used primarily for evaluating individuals for clinical or psychoeducational purposes, such as for the assessment of MENTAL RETARDATION, GIFTEDNESS, LEARNING DISABILITY, and brain injury. Many of the current tests of mental ability provide a general overall score (usually referred to as an INTELLIGENCE QUOTIENT [IQ] or some similar designation) and a profile of mental abilities. This profile is particularly useful for evaluating the individual's cognitive strengths and weaknesses. Most of the tests reviewed in this article can be administered in approximately 1 hour, although in some cases administration time may be longer.

Individually administered tests of intelligence, such as the Wechsler and Stanford-Binet scales, have several advantages over group tests, such as the SCHOLASTIC ASSESSMENT TEST (SAT). First, group-administered tests are limited in the range of behaviors they are able to

assess. The tests usually are in a multiple-choice format, thereby limiting the individual's responses to predetermined categories. Second, group-administered tests do not permit the examiner to obtain any information about the strategies that the examinees use to solve problems. Finally, such tests do not allow examiners to observe how the individual solves problems, or to probe for further information or clarification of responses.

Individually administered tests, on the other hand, are more flexible than group-administered tests. An examinee's strategies, reaction time, work habits, ability to profit from cues, language and motor skills, and attitude and motivation can all be monitored. This information, coupled with the examinee's responses to test questions, is valuable for understanding the examinee's level of cognitive ability.

WECHSLER ADULT INTELLIGENCE SCALE—REVISED (WAIS—R)

The WAIS—R (Wechsler, 1981) is a comprehensive battery of eleven subtests grouped into Verbal and Performance sections. The WAIS—R covers an age range from 16 years, 0 months, to 74 years, 11 months, and is designed to measure both verbal and nonverbal aspects of intelligence. Its original version, the Wechsler-Bellevue Intelligence Scale, was first introduced in 1939 by David WECHSLER.

A brief description of the six Verbal Scale subtests follows:

1. Information: Examinees answer questions that sample a broad range of general knowledge.
2. Digit Span: Examinees recall digits in a forward order and then in a backward order.
3. Vocabulary: Examinees define words.
4. Arithmetic: Examinees solve arithmetic problems.
5. Comprehension: Examinees answer questions dealing with a range of situations and problems.
6. Similarities: Examinees explain the similarity between a pair of words.

The five Performance Scale subtests are as follows:

7. Picture Completion: Examinees identify the missing element in drawings of common objects.
8. Picture Arrangement: Examinees arrange a series of pictures into a logical sequence.

9. Block Design: Examinees reproduce two-dimensional designs using blocks.
10. Object Assembly: Examinees assemble jigsaw-puzzle pieces to form common objects.
11. Digit Symbol: Examinees copy symbols following a key.

For each subtest a raw score is obtained (individual correct responses receive varying point values). The raw scores for each subtest are then converted into standard scores ($M = 10$, $SD = 3$). These standard scores are summed and converted into a Verbal IQ, Performance IQ, and Full Scale IQ ($M = 100$, $SD = 15$).

The WAIS—R is well-standardized, including 1,880 people in the standardization sample. The WAIS—R provides high internal consistency and test-retest reliabilities for the three IQs. The Verbal Scale IQ, however, is more stable than the Performance Scale IQ. Reliabilities for individual subtests are less adequate. The WAIS—R has adequate concurrent and construct validity.

Strengths of the WAIS—R are its RELIABILITY and VALIDITY, good standardization, and division into Verbal and Performance components, which has proved helpful in clinical and psychoeducational applications and facilitates the assessment of sensory-impaired individuals (such as the blind or the deaf). In addition, there are good administration procedures and guidelines. Limitations of the WAIS—R are its limited capacity to differentiate individuals with extremely high or extremely low abilities, nonuniform ranges of scaled scores for subtests, ambiguity in scoring some verbal responses, and failure to identify rationale for establishing discontinuance criteria. Nevertheless, the WAIS—R is a popular and widely used measure and has proved to be useful in both clinical and research applications. (See also WAIS—R SUBTESTS.)

WECHSLER INTELLIGENCE SCALE FOR CHILDREN—THIRD EDITION (WISC—III)

The WISC—III (Wechsler, 1991), another of the Wechsler tests, is a comprehensive battery of thirteen subtests that cover the age range of 6 through 16 years. It measures both verbal and nonverbal aspects of intelligence. The WISC—III is similar to its prede-

cessor, the WISC–R, which has been one of the most popular measures for child intellectual assessment. The differences in the WISC–III include the addition of a new subtest, Symbol Search, improvements in the quality of pictorial material, a revised order of administering the subtests, and slight modifications in scoring and administration guidelines.

Like the WAIS–R, subtests on the WISC–R are grouped into Verbal and Performance sections. The six Verbal Scale subtests are Information, Similarities, Arithmetic, Vocabulary, Comprehension, and Digit Span (optional). The seven Performance Scale subtests are Picture Completion, Picture Arrangement, Block Design, Object Assembly, Coding, Mazes (optional), and Symbol Search (optional). The WISC–III subtests that also are on the WAIS–R share the general features described above. Therefore, only those subtests that differ from the WAIS–R are explained below.

The three subtests unique to the WISC–III are as follows:

1. Coding: Examinees copy symbols according to a specified pattern as quickly as possible.
2. Mazes: Examinees complete a series of mazes with pencil and paper.
3. Symbol Search: Examinees look at a symbol(s) and decide whether it appears in an array of symbols.

The WISC–III follows a procedure similar to that described for the WAIS–R for scoring items and converting raw scores to scaled scores and scaled scores to IQs. The WISC–III also provides four Index Scores: Verbal Comprehension, Perceptual Organization, Freedom from Distractibility, and Processing Speed. The index scores are essentially factor scores (based on factor analysis), each of which is based upon a different combination of subtests.

The WISC–III is a well-standardized test, covering 2,200 children in the standardization sample. The WISC–III has excellent internal-consistency reliability for the three IQs and acceptable internal-consistency reliability for the subtests. Test–retest reliability, however, is less adequate, particularly for the Performance Scale IQ and the Performance Scale subtests. In addition, practice effects are larger on the Performance Scale (an average of 12 points) than on the Verbal Scale (an average of about 2 points) over a 12-to-63-day

retest period. Concurrent validity of the WISC–III is high.

The strengths and weaknesses of the WISC–III are similar to those described for the WAIS–R. Because it is a relatively new revision, however, research is needed to determine how its scores compare to the scores of the prior edition and other intelligence tests.

WECHSLER PRESCHOOL AND PRIMARY SCALE OF INTELLIGENCE– REVISED (WPPSI–R)

The WPPSI–R (Wechsler, 1989) is designed to evaluate verbal and nonverbal intelligence of children aged 3 years, 0 months, through 7 years, 3 months. The first edition was published in 1967. Its 1989 revision primarily entailed updating the norms, improving the test content so as to appeal more to young children, adding a new subtest (Object Assembly), and expanding the age range of the original WPPSI.

Like the other Wechsler tests, the twelve subtests are divided into Verbal and Performance sections, with six subtests comprising the Verbal Scale (Information, Comprehension, Arithmetic, Vocabulary, Similarities, Sentences), and six comprising the Performance Scale (Object Assembly, Geometric Design, Block Design, Mazes, Picture Completion, Animal Pegs). As many of the subtests are similar to those on the WAIS–R and WISC–III, only those subtests that differ will be described below.

1. Sentences: Examinees repeat sentences of increasing complexity.
2. Geometric Design: Examinees select the matching design from four choices (earlier items), and they copy geometric designs using pencil and paper (later items).
3. Animal Pegs: Examinees place colored pegs in corresponding holes on a board.

The computation of scores is essentially the same as for the WAIS–R and WISC–III.

The WPPSI–R is a well-standardized test, based on scores of 1,700 children in its normative sample. The IQs provided by the WPPSI–R are highly reliable, except at age seven, when it is recommended that the WISC–III be administered instead. Some of the subtests have low reliability for some ages, although many

are adequate. Test-retest reliability is generally strong, but there is evidence of practice effects on performance subtests. Studies of concurrent validity yielded generally adequate correlations with other measures of children's intelligence.

Strengths of the WPPSI are its good standardization, strong reliability and validity, good administrative procedures and manual, varied test materials, and useful diagnostic information. Limitations include the low reliability of individual subtests for some ages, limited capacity to differentiate abilities at the lower end of the scale, nonuniformity of subtest scores, long administration time, and overlap of some items with the WISC-III, making it less than a completely independent test. Nevertheless, the WPPSI-R is a valuable test for assessing the cognitive ability of young children.

STANFORD-BINET INTELLIGENCE SCALE, FOURTH EDITION (SB IV)

The SB IV (Thorndike, Hagen, & Sattler, 1986) is a battery of fifteen subtests that measure both verbal and nonverbal components of intelligence. The test, which covers an age range of 2 through 23 years, represents the latest edition of the Stanford-Binet Intelligence Scale. It has its roots in the earliest intelligence tests developed at the beginning of the twentieth century by Alfred BINET, Théophile Simon, and Lewis TERMAN. The SB IV differs in many ways from its predecessors, including its theoretical basis and major revisions of the way items are administered.

There are four areas assessed by the SB IV: verbal reasoning, quantitative reasoning, abstract/visual reasoning, and short-term memory. The subtests each contribute to one of the area scores and are described briefly below.

1. Vocabulary: Examinees name pictured objects on the first 14 items and define the words on the remaining items.
2. Bead Memory: Examinees reproduce bead patterns by identifying them in photographs or by placing beads on a stick.
3. Quantitative: Examinees solve quantitative, arithmetic-like problems.
4. Memory for Sentences: Examinees repeat successively longer sentences.
5. Pattern Analysis: Examinees place pieces in a

form board for the first six items and reproduce two-dimensional designs using blocks for the remaining items.

6. Comprehension: Examinees answer questions dealing with social comprehension.
7. Absurdities: Examinees identify the incongruity in pictures.
8. Memory for Digits: Examinees recall digits presented in a forward direction and in a backward direction.
9. Copying: Examinees reproduce designs with blocks or copy geometric designs with paper and pencil.
10. Memory for Objects: Examinees recall pictured objects in the exact sequence in which they were presented by pointing to the pictured items on a card.
11. Matrices: Examinees select the object, design, or letter that best completes the matrix.
12. Number Series: Examinees predict the next two numbers in a series.
13. Paper Folding and Cutting: Examinees select the picture that shows how a folded and cut piece of paper would look unfolded.
14. Verbal Relations: Examinees indicate how the first three items are alike but different from the fourth.
15. Equation Building: Examinees arrange numbers and mathematical signs into an equation.

For each subtest a raw score is obtained (items receive 1 point if correct and 0 points if incorrect). The raw scores are then converted into three types of standard scores: standard age scores (or scaled scores) for the subtests ($M = 50$, $SD = 8$), area scores ($M = 100$, $SD = 16$), and a Composite Score (analogous to the IQ; $M = 100$, $SD = 16$).

The SB IV is a well-standardized test, including over 5,000 people in the normative sample. Internal-consistency reliability of the Composite Score is very high, and all subtests are reliable. With regard to test-retest reliability, the Composite Score has demonstrated adequate stability for normal populations, more so than individual subtest scores. The SB IV also has adequate concurrent validity, demonstrated by correlations with several measures of ability.

The strengths of the SB IV are its reliability and validity, excellent standardization, good administration

procedures, adequate administrative guidelines and test materials, and helpful scoring criteria. Limitations of the SB IV are its overly long administration time, lack of a single, comparable battery throughout the age ranges covered by the scale (only six subtests are given to all ages), variable range of scores for different ages, and limited factor analytic support for the area scores. Nevertheless, the SB IV is a potentially powerful tool for assessment of cognitive ability, as it represents an exceptional collection of individual measures.

KAUFMAN ASSESSMENT BATTERY FOR CHILDREN (K-ABC)

The K-ABC (Kaufman & Kaufman, 1983) is a measure of intelligence and achievement designed for children aged 2 years, 6 months, through 12 years, 5 months. The K-ABC was designed to provide clinicians with an alternative method of evaluating intelligence. Four scales are included in the battery: Sequential Processing Scale, Simultaneous Processing Scale, Achievement Scale, and Nonverbal Scale.

The Sequential and Simultaneous Processing Scales are hypothesized to reflect the child's style of problem solving and information processing. Scores from these two scales are combined to form the Mental Processing Composite, which serves as the primary measure of intelligence on the K-ABC. The Sequential and Simultaneous Processing Scales were designed to reduce the effects of verbal processing and ethnic bias. In contrast, the Achievement Scale is more heavily loaded on verbal skills. The Nonverbal Scale is not a unique scale; it is composed of those subtests from the other scales that are nonverbal. Not all subtests are administered at every age; of the 16 subtests in the battery, no more than thirteen are administered to any one child. The subtests are described briefly below.

1. Magic Window: Examinees identify a picture that the examiner moves past a narrow slit (making the picture only partially visible).
2. Face Recognition: Examinees select from a group photograph the one or two faces that were shown in a preceding photograph.
3. Hand Movements: Examinees imitate a series of hand movements.
4. Gestalt Closure: Examinees name an object or

scene pictured in a partially completed "inkblot" drawing.

5. Number Recall: Examinees repeat a series of digits.
6. Triangles: Examinees reproduce two-dimensional designs using blocks.
7. Word Order: Examinees touch a series of pictures in the same sequence as they were named by the examiner.
8. Matrix Analogies: Examinees select the picture or figure that completes the matrix.
9. Spatial Memory: Examinees recall the placement of pictures on a page that was exposed for a brief interval.
10. Photo Series: Examinees place photographs of an event in chronological order.
11. Expressive Vocabulary: Examinees name objects pictured in photographs.
12. Faces and Places: Examinees name a well-known person or place from a picture.
13. Arithmetic: Examinees answer mathematical questions.
14. Riddles: Examinees name an object or concept described by several of its characteristics.
15. Reading/Decoding: The task involves word recognition—reading words out of context.
16. Reading/Understanding: Examinees act out commands given in words or sentences.

Raw scores for the global scales are converted into standard scores ($M = 100$, $SD = 15$). Raw scores from the subtests are converted into scaled scores ($M = 10$, $SD = 3$). The primary index of mental ability, the Mental Processing Composite, is then derived from scores on the Simultaneous and Sequential Processing Scales.

The K-ABC standardization was adequate, including 2,000 children in the standardization sample. Reliability for the Mental Processing Composite and the Achievement Scale is satisfactory, as is test-retest reliability. Construct, concurrent, and predictive validity have also been demonstrated.

The K-ABC is a unique contribution to the assessment scene. It has generated much controversy, however, including serious questions about its theoretical rationale, factor structure, and proposed remediation strategies. For example, there is no evidence that the

subtests purported to measure separate and independent processes actually do so. The lack of verbal comprehension or reasoning items on the Mental Processing Composite, along with the heavy reliance on short-term memory are fundamental weaknesses of the test. The K-ABC also has limited capacity to differentiate abilities at the upper and lower ends of the scale. Finally, the use of the term "mental processing" for some subtests and the term "achievement" for others is potentially misleading, since all the subtests involve processing. Nevertheless, the K-ABC may prove to be useful in certain situations, particularly when information is needed about nonverbal cognitive abilities.

McCARTHY SCALES OF CHILDREN'S ABILITIES

This test (McCarthy, 1972) is a well-standardized and psychometrically sound measure of the cognitive abilities of young children. It covers the age range of 2 years, 6 months, to 8 years, 6 months. The McCarthy Scales render a general measure of intellectual functioning called the General Cognitive Index (GCI), as well as a profile of abilities that includes measures of verbal ability, nonverbal reasoning ability, number aptitude, short-term memory, and coordination.

The McCarthy Scales include the following six scales: Verbal Scale, Perceptual-Performance Scale, Quantitative Scale, Memory Scale, Motor Scale, and General Cognitive Scale. Each scale contains three to fifteen subtests. Some of the subtests are used on more than one scale. Five scale indices (standard scores; $M = 50$, $SD = 10$) are derived from the McCarthy Scales subtests. The overall GCI ($M = 100$, $SD = 16$) is similar to an IQ score.

The standardization of the McCarthy Scales is excellent, covering 1,032 children in the standardization sample. Reliability, both split-half and test-retest, are adequate for the scale. Regarding validity, although concurrent validity coefficients are satisfactory, the GCI has been shown not to be interchangeable with the IQ scores generated from other measures of intelligence. Predictive validity is satisfactory, although construct validity appears to be questionable. For this reason, the scales may be better measures of general

cognitive abilities than of the specific abilities designated by the names of the scales.

Although the McCarthy Scales have many strengths, it also has many weaknesses. Because it does not provide scores that are equivalent to those provided by the WISC-III, caution should be exercised in using GCIs for placement decisions, especially in the assessment of mentally retarded or gifted children. The lack of standard scores for each test by age level also limits the diagnostic usefulness of the scale. In spite of these limitations, however, the McCarthy Scales provide a profile of abilities that may be quite useful in evaluating young children. The test is useful for assessing the cognitive ability and, to a lesser extent, the motor abilities of young children and therefore deserves serious consideration.

BAYLEY SCALES OF INFANT DEVELOPMENT—SECOND EDITION (1993)

This is a carefully produced measure of infant development for children from 1 to 42 months of age. Two standard scores are provided: a Mental Developmental Index (MDI), obtained from the Mental Scale, and a Psychomotor Development Index (PDI), obtained from the Motor Scale. A Behavior Rating Scale that assesses the infant's and child's attitude, interest, emotion, energy, activity, and responsiveness also can be completed by the examiner.

The Mental Scale contains 178 items arranged by one-month intervals. Items involve sensory-perceptual acuity, discriminations, memory, learning and problem solving, verbal ability, and concept formation. The 111 items in the Motor Scale assess muscle control (control of the body) and gross and fine motor abilities, such as sitting, standing, walking, and grasping. Raw scores on the Bayley Scales-II are converted into the MDI and PDI, which are normalized standard scores ($M = 100$, $SD = 15$).

The Bayley Scales-II were standardized on a representative national sample of 1,700 normal infants and children. Reliability and validity meet acceptable standards. The test significantly correlates with the first edition of the scale and with other measures of intelligence, such as the WPPSI-R.

The Bayley Scales–II require considerable practice and experience to administer. The test is an excellent addition to the area of infant assessment and is by far the best measure of infant development. (See also Nancy BAYLEY.)

WOODCOCK-JOHNSON TESTS OF COGNITIVE ABILITY

The Woodcock-Johnson Tests of Cognitive Ability (Woodcock & Johnson, 1989) measure both verbal and nonverbal intelligence. It is a revision of the Woodcock-Johnson Psycho-Educational Battery, a comprehensive assessment measure that includes the Woodcock-Johnson Tests of Cognitive Ability, Tests of Achievement (measures several achievement areas, including reading, spelling, capitalization, and punctuation), and Scales of Independent Behavior (measures several adaptive behavioral skills, including motor skills, social interaction and communication skills, and personal and community living skills). The revised Tests of Cognitive Ability contain twenty-one subtests assessing the domains of Visual Processing, Auditory Processing, Processing Speed, Short-Term Memory, Long-Term Retrieval, Comprehension Knowledge, Fluid Reasoning, and Quantitative Ability; seven subtests are standard and fourteen are supplemental. The battery is designed for ages 2 through 90+. The revised battery represents an improvement over its predecessor by adhering closely to its theoretical model; this allows for more confidence in interpreting results.

The Woodcock-Johnson yields a Broad Cognitive Ability composite score ($M = 100$, $SD = 15$), which is analogous to an IQ score, and the composite score for preschool-aged children is termed Early Development. The test also generates individual subtest scores and cluster scores ($M = 100$, $SD = 15$). Grade-level scores are provided for subtests and cluster scores, and a number of other scoring options are available (such as t -scores, normal curve equivalents, stanines, and a Relative Mastery Index).

The Woodcock-Johnson is well standardized, covering 6,359 persons in its normative group. Reliability appears to be satisfactory, although the overall composite and cluster scores are more reliable than the subtest scores. To date, not much is known about the

stability of scores on the Woodcock-Johnson. Correlations with other tests of intelligence are moderate, and lower than one would expect for a multisubtest measure of intelligence.

Because it is a relatively new measure, more research is needed on the revised Woodcock-Johnson Tests of Cognitive Ability before its psychoeducational use can be determined. For example, the tests appear to be heavily weighted in favor of subtests that measure academic achievement skills, and there are few items per subtest relative to the large age range sampled. In general, although the Woodcock-Johnson Psycho-Educational Battery has some unique features not found in other cognitive and achievement tests, the Cognitive Ability Full Scale score should not be used as a replacement for other standardized measures of intelligence, such as the WISC–III or SB IV.

OTHER INDIVIDUALLY ADMINISTERED TESTS OF INTELLIGENCE

Other measures used to assess intelligence are the Battelle Developmental Inventory (BDI), Columbia Mental Maturity Scale, Denver Developmental Screening Test–Revised, Detroit Tests of Learning Aptitude (DTLA), Differential Abilities Scale (DAS), Extended Merrill Palmer Scale, Goodenough-Harris Drawing Test, Hiskey-Nebraska Test of Learning Aptitude (HNTLA), Leiter International Performance Scale, Peabody Picture Vocabulary Test–Revised (PPVT–R), Raven Progressive Matrices, and Test of Nonverbal Intelligence–2 (TONI). These tests are less frequently used, but are useful in certain situations, such as for screening purposes.

COMMENT ON INDIVIDUAL TESTS

All the tests reviewed in this section are valuable for the assessment of individual mental ability. Each has its strengths and weaknesses, but they all make a contribution to our understanding of intelligence. Because of the need to administer them in a standardized manner, however, they tend to penalize individuals who work in a slow, deliberate manner or who offer unique, unconventional responses. In addition, when individual tests are repeatedly administered to the

same individual, practice effects can occur. Practice effects are difficult to interpret because it is not clear whether they reflect genuine improvement or are simply the result of having seen or experienced the items previously. In summary, individual tests of mental ability form the cornerstone of most assessment batteries in the field of clinical and psychoeducational assessment.

(See also: KAUFMAN ASSESSMENT BATTERY FOR CHILDREN; MCCARTHY SCALES OF CHILDREN'S ABILITIES; WECHSLER SCALES OF INTELLIGENCE; WOODCOCK-JOHNSON TESTS OF COGNITIVE ABILITY.)

BIBLIOGRAPHY

- BAYLEY, N. (1993). *Bayley Scales of Infant Development—Second Edition*. San Antonio, TX: The Psychological Corporation.
- KAUFMAN, A. S., & KAUFMAN, N. L. (1983). *K-ABC: Kaufman Assessment Battery for Children*. Circle Pines, MN: American Guidance Service.
- MCCARTHY, D. (1972). *McCarthy Scales of Children's Abilities*. New York: The Psychological Corporation.
- THORNDIKE, R. L., HAGEN, E. P., & SATTLER, J. M. (1986). *Technical manual, Stanford-Binet Intelligence Scale: Fourth Edition*. Chicago: Riverside Publishing.
- WECHSLER, D. (1981). *Manual for the Wechsler Adult Intelligence Scale—Revised (WAIS-R)*. New York: The Psychological Corporation.
- WECHSLER, D. (1989). *Manual for the Wechsler Preschool and Primary Scale of Intelligence—Revised (WPPSI-R)*. New York: The Psychological Corporation.
- WECHSLER, D. (1991). *Manual for the Wechsler Intelligence Scale for Children—Third Edition*. New York: The Psychological Corporation.
- WOODCOCK, R. W., & JOHNSON, M. B. (1989). *Woodcock-Johnson Psycho-Educational Battery—Revised*. Allen, TX: DLM Teaching Resources.

JEROME M. SATTLER
 NAOMI G. SINGER

INFANCY Very young infants have an active mental life. They are constantly learning and developing new ideas. They must, in order to survive and succeed. However, children differ in their mental abil-

ities, just as they do in their temperament and emotions, and they vary in the experiences they have as well as in what they learn in their interactions with the social and physical environment. Both experience and ability influence mental development, for what children learn and the capacities they demonstrate in doing so project to later life. This perspective reflects the significance of infant cognition and early experience on the ultimate development of intelligence.

Consider what kind of mental capacity you might engineer for a baby. Certainly, you would not want to fix a baby's intelligence in advance of the child's actually experiencing the world. It would be shortsighted and counterproductive to eliminate potentially important influences of experience. Intelligence includes the ability to adapt successfully to the environment. By the same token, however, you would probably not want to leave mental development wholly to experience. Experiences among individuals as they are growing up can vary haphazardly. Finally, you would probably want to endow a baby with some tools to cope with and take advantage of those experiences. In fact, all these "design criteria" appear to have been met in the evolution of infant cognition.

TRADITIONAL APPROACHES TO EVALUATING COGNITION IN INFANCY

Intelligence testing dates only from the beginning of the twentieth century (Carroll, 1986). Soon after this tradition took root, two compelling questions related to the earliest manifestations and development of intelligence naturally arose: Can "intelligence" in infancy be measured? What does "intelligence" in infancy foretell about "intelligence" later in life?

Many early-twentieth-century investigators developed groups of standardized tasks for infants that were graded in difficulty by age, and for years their sequences, scales, and schedules of infant behavior proved valuable in defining normative development and in identifying children at risk for developmental delay. Perhaps the best known and most widely adopted of such tests has been the Bayley Scales of Infant Development (Bayley, 1969). The Bayley Scales assess motor capacity, sensation, perception, cognition, memory, language, and social behavior in infants over the first 2½ years of life.

The validity of such infant tests has typically been assessed by comparing infants' performance early in life with their performance years later as children or even adults. The presumption of this strategy is that if individuals who perform well on infant tests later do well on standardized intelligence tests, then the infant tests must be revealing of something about "intelligence" in infancy. (Of course, what exactly that something is itself is still not known.)

Nancy BAYLEY's classic longitudinal study (1949) exemplifies both the findings and the far-reaching conclusions that characterize much of the early tradition in infant testing. Bayley followed children from 3 months to 18 years of age, correlating their scores on her test in infancy and in early childhood with their intelligence test scores in young adulthood. Essentially she found no association between test performance in infancy and intelligence test performance in maturity. Only after children reached about 5 to 7 years of age did the association between child scores and eventual adult scores rise, subsequently attaining a very high level between 11 and 18 years. Her findings have been replicated many times (Kopp & McCall, 1980).

The lack of predictive validity in traditional infant tests led to several profound general conclusions about human mental growth from infancy. Some theorists argued that there is no general intelligence factor in mental life; otherwise, smarter infants would be smarter children. Some argued that if such a factor exists, it is not fixed or stable in individuals, at least across the early part of the life span. Some argued that mental growth progressively changes so that intelligence in infancy is very different from mature intelligence. These interpretations of the findings about mental development differ considerably, but all at base contend that infant mental life is disconnected from mature mental life, and all were grounded in data showing the lack of cross-age predictability in intelligence from traditional infant tests.

Did nature design infant mental life to have no continuing relation to later development? The lack of stability in mental development from infancy has been challenged on several grounds. First, traditional infant tests are not themselves very reliable; that is, infant performance at one time does not accurately predict infant performance just a short time later. (Reliability means maintenance of rank-order performance of in-

dividuals over time; a test should achieve adequate reliability to be predictive.) Second, the kinds of items put to young infants in most traditional infant tests tap largely sensory, motor, and affective responses, such as orienting, reaching, and smiling. For an older child, very different items are used in evaluating intelligence—normally, skills related to language, reasoning, and memory. Thus, the constructs compared across age bear little or no conceptual relation to one another in the first place.

If the hypothesis of stability in mental development from infancy is going to be tested, then, longitudinal assessments of predictability in the cognitive domain should evaluate infant skills in ways that are psychometrically sound, that are relatively free of motor requirements and affective components, and that conceptually parallel or subserve cognitive functions typical of childhood. Information-processing measures in infancy that meet these criteria (notably habituation) have been identified. These measures show individual variation and possess adequate reliability. Performance on them predicts cognitive competencies in childhood—not perfectly, but better than do scores on traditional infant tests. As a consequence, they have meaningful implications for understanding mental development generally.

INFANT COGNITION

An information-processing perspective on cognitive development focuses on the activities and mechanisms used to bring information in the environment into the cognitive system. Representing and manipulating information mentally implies discrimination, encoding, storage, and retrieval and recognition. Information processing is not a comprehensive theory of cognition or of cognitive development, but rather an approach to studying mental processes (Kuhn, 1992).

Infant Information Processing. How is information processing in infancy studied? There are now several ways. Consider the following: an infant will typically orient and attend to a new stimulus in the environment, but if that stimulus remains available to view, the infant's attention to it will usually diminish. Habituation is the decline in attention to a stimulus that is available continuously or is presented repeatedly. Habituation is thought to reflect mental proc-

esses concerned with the ongoing construction of some sort of central mental representation of stimulus material and the comparison between stimulation being presented and that mental representation. Habituation assesses the infant's reaction to an aspect of the environment that is unchanging and represents an elementary kind of learning (Bornstein, 1985).

Different strands of data help to confirm an information-processing interpretation of habituation. As they grow and develop, infants process information more and more efficiently. Across the first year of life, for example, infants show a steady decline in the amount of time they require to inspect a stimulus in order to reach a constant habituation criterion (Bornstein, Pêcheux, & Lécuyer, 1988; Fantz, 1964). Relatedly, normally developing babies habituate more efficiently than babies born at risk for cognitive delay. Infants of a given age and mental status also require more time to encode information in a complex stimulus than in a simple one (Caron & Caron, 1969). Further, infants habituated to one stimulus can later distinguish a new stimulus in comparison with their internal representation of the familiar one; the infants' novelty responsiveness to stimulus change provides clear and converging support for a representation-making interpretation of habituation. Indeed, even newborn babies habituated to a stimulus viewed through only one eye later recover attention when looking at a novel stimulus (in comparison with the habituation stimulus) through the other eye (Slater, Morison, & Rose, 1983). This "interocular transfer" indicates that some information about environmental stimulation is processed centrally in the brain from the beginning of life.

Habituation is not an idiosyncratic laboratory phenomenon, but is typical of infants' everyday interactions with people and objects in the world. Moreover, characteristics of habituation at home are similar to those measured more formally in the laboratory (Bornstein & Ludemann, 1989).

Individual Variation and Reliability in Infant Information Processing. Considerable research has been conducted on two important measurement concerns with respect to infant information processing—namely, distinguishing variation among individual babies, and determining whether individual differences constitute reliable, that is, repeatable, char-

acteristics of babies. If individual differences exist in babies, and those differences are reliable in the sense of being a part of the infant, we can expect that they will play a meaningful role in development.

Infants differ among themselves considerably in habituating, that is, they show individual variation with respect to the rate, amount, and pattern of attention decrement (Bornstein, 1989a). Whether measured quantitatively or qualitatively, individual differences in habituation also show moderate reliability. That is, given the same testing conditions, a baby is likely to habituate in relatively the same way on different occasions if the latter are spaced reasonably closely together in time (Bornstein, 1989a; Colombo & Mitchell, 1990). Of course, differences in procedures, as well as other sources of unreliability, such as infant age and state and the nature of the stimuli, can affect these outcomes. Variation and reliability are important because they say that, at least to some degree, infants differ in information-processing capacity, and that those differences are in the infant, rather than in the procedure, the stimulus, or the situation.

Predictive Validity in Infant Information Processing. Generally speaking, infants who process information more efficiently are thought to acquire knowledge of a stimulus more quickly and thoroughly. Thus, greater decrements, quicker decays, and shorter total looking times in habituation are generally considered to index more efficient information processing (Bornstein, 1985, 1989a; Colombo & Mitchell, 1990; Slater, 1990). We certainly understand the everyday meaning of "quick" and "slow" when used to connote a person's intelligence. Learning rate and reactivity are traditional parts of our definition of intelligence, and inspection time is a part of intelligence test performance (Deary, 1992). Attention has long been considered to be a basic component of mental life, and attention has traditionally been viewed as a key factor of intelligence (Cooper & Regan, 1982).

Measures of habituation in infancy and childhood have proved to possess some validity in relation to intelligent performance. Infants and young children who habituate efficiently tend also, at about the same point in their young lives, to prefer complex patterns over simple ones, to show advanced sensorimotor development, to explore the environment rapidly, to play in relatively sophisticated ways, to show better rec-

ognition memory, to solve problems quickly and to attain concepts readily, and to excel at oddity identification, picture matching, and block configuration (Bornstein, 1985, 1989a).

How well does habituation in infancy predict childhood intelligence test performance? Measures of decrement of attention in infancy have been found to possess moderate predictive validity (Bornstein, 1985, 1989a; McCall & Carriger, 1993). Infants who habituate efficiently in the first 6 months of life will, between 2 and 12 years of age, perform better on traditional assessments of cognitive competence, including measures of both psychometric intelligence and representational competence, such as language and play. Significantly, such predictive validity has been observed in different laboratories, with different populations of both normal and at-risk infants, for different measures in infancy and in childhood, and across different modalities, including visual and auditory. The prediction between information-processing measures in infancy and cognitive performance in childhood is significantly higher than that between traditional infant tests and childhood intelligence tests, and they have a mean correlation of about .50.

Habituation is not the only cognitive process identified in infants. Infants possess a small armamentarium of information-processing abilities. For example, if an infant sees the familiar alongside a novel stimulus after habituation, the infant's looking at the familiar stimulus is usually depressed relative to attention displayed to the new stimulus. Attention to the new stimulus is called responsiveness to novelty. Novelty responsiveness is typical of infants' everyday interactions with objects and people in the world as well (Fagan, 1992). Individual variation in novelty responsiveness possesses predictive validity for childhood intelligence test performance (Bornstein, 1989a; Fagan, 1992; McCall & Carriger, 1993).

Of course, information-processing capacities in infancy could covary with some other factor(s) in the child that may be responsible for the predictive association between habituation or novelty responsiveness in infancy and childhood intelligence. Besides intelligence, it presumably takes being developed perceptually, possessing a vigilant or persistent cognitive style, and having an attentive temperament to perform well first on the infant tests and later on an intelligence

test. These arguments lose force, however, in light of the fact that infant performance on the Bayley Scales (as well as on other traditional infant tests) shows no predictive validity. That is, there is no reason that performance on traditional infant tests should not share variance with motivation, personality, or whatever, in the same way information-processing capacities would. Therefore, if these generalized factors were carrying the association, they should do so equivalently for traditional infant test performance and information-processing capacities.

In the information-processing approach to infant cognition, researchers have developed techniques that permit some access to the mind of the young child. Babies habituate attention to familiar stimuli and recover attention to novel ones. They vary in the ways they do, and these variations show reasonable reliability—that is, they are a part of the child. Further, infants who perform well on information-processing tasks later as children tend to perform well on intelligence tests, measures of language ability, and the like. The speed, accuracy, and completeness of encoding a stimulus into memory, which are presumably measured by habituation and novelty responsiveness, appear to relate to the vocabulary, abstract abilities, and memory skills assessed by childhood tests of intelligence.

Although the data on stability in mental development are telling, it is important to underscore several facts. Much is known about information processes, but it is also fair to say that much is not known about them. For example, whether individual variation in habituation and novelty responsiveness reflects central nervous system integrity, maturation, or function, or whether it reflects strategy differences in information processing, is not known. Further, the infant information-processing measures are still far from perfect predictors of childhood performance. Finally, stability certainly does not mean that intelligence is innate, or that intelligence is fixed in early life, or that the infant is alone in developing intelligence. Findings of stability do, however, reinforce the view that cognition in infancy is likely to play a meaningful role in later development. Notably, the verbal and logical kinds of intelligence studied so far as predicted outcomes in childhood themselves relate to performance in many domains of theoretical and practical significance in

everyday life, such as academic attainment and occupational achievement. Intelligence is a multifactorial psychological construct; this means that there are many kinds of intelligence to be studied. It may be that these information-processing capacities in infancy project to one or more of them, beyond simple IQ.

INFANT MENTAL DEVELOPMENT IN ITS SOCIAL CONTEXT

Modern behavior genetics argues that reliable individual differences in intelligence can be expected to reflect biological inheritance to some degree (Bouchard, Lykken, McGue, Segal, & Tellegen, 1990; Plomin, 1990). But experience in the world is either a principal source of individual cognitive growth or a major contributing component (Dixon & Lerner, 1992; Kuhn, 1992). For example, comparative studies of mental growth in children from adoptive and non-adoptive families confirm the view that aspects of intelligence are shared biologically, but also that identifiable aspects of the mother–infant caregiving environment predict developmental status in the child (Plomin, 1990). Further, increasing evidence indicates that specific parental activities relate concurrently and predictively to specific aspects of child performance in specific ways, and that parent and infant together influence the course and outcome of the child's mental development (Bornstein, 1989b; Wachs, 1992). Defining which aspects best predict which components of developed cognition warrants intense empirical focus on everyday activities in the family.

It is critical to note in this connection that habituation in infancy predicts cognitive status in childhood independent of effective family influences (such as maternal behaviors). This suggests an unmediated (direct) tie between infant information processing and childhood cognitive skills. In other words, some stability appears to be in the individual. Relative to experience, it could also be that babies who process information efficiently expose themselves to more optimal amounts, kinds, or patterns of environmental stimulation. Significantly, too, infants who habituate more efficiently have mothers who treat them in more cognitively sophisticated ways later in childhood. Clearly, infants and caregivers *jointly* contribute to the developing cognitive competencies of children.

CONCLUSIONS

The job of the mind in development is to coordinate with physical and social reality. The assessment of cognitive capacity and potential early in life, the determination of endogenous and exogenous factors in their origins and expression, and the investigation of the predictive validity of such factors for childhood cognitive performance are abiding and compelling topics in infancy studies. One productive approach to evaluating mental life in infancy is associated with information processing. This school of thought attempts to define antecedents of mature cognition and focuses on the ABC of how children learn, acquire, and mentally manipulate information. A complementary approach centers on identifying influences of specific experiences on child mental life.

Three revolutions have taken place in the study of early cognitive development. Where before the thrust of most investigations centered on the general developmental function, contemporary studies have profitably revisited the question of individual variation. Where before traditional infant assessments showed little predictive validity for later cognitive stature, contemporary studies using information-processing techniques have unearthed significant levels of stability in cognitive performance from infancy. And, where before researchers suspected that early experiences ought to influence formative intellectual growth, but failed in efforts to identify specific mechanisms, contemporary studies focusing on newly operationalized variables have begun to unearth significant and specific predictive relations. Together, these new points of view meet the design criteria for mental development suggested at the start of this article, and they show that endogenous and exogenous forces interact to coordinate individual mind with external reality.

(See also: INFANT TESTS AS MEASURES OF EARLY COMPETENCE.)

BIBLIOGRAPHY

- BAYLEY, N. (1949). Consistency and variability in the growth of intelligence from birth to eighteen years. *Journal of Genetic Psychology*, 75, 165–196.
- BAYLEY, N. (1969). *Bayley scales of infant development*. New York: The Psychological Corporation.

- BORNSTEIN, M. H. (1985). Habituation of attention as a measure of visual information processing in human infants: Summary, systematization, and synthesis. In G. Gottlieb & N. A. Krasnegor (Eds.), *Measurement of audition and vision in the first year of postnatal life: A methodological overview*. Norwood, NJ: Ablex.
- BORNSTEIN, M. H. (1989a). Stability in early mental development: From attention and information processing in infancy to language and cognition in childhood. In M. H. Bornstein & N. A. Krasnegor (Eds.), *Stability and continuity in mental development: Behavioral and biological perspectives*. Hillsdale, NJ: Erlbaum.
- BORNSTEIN, M. H. (1989b). Between caretakers and their young: Two modes of interaction and their consequences for cognitive growth. In M. H. Bornstein & J. S. Bruner (Eds.), *Interaction in human development*. Hillsdale, NJ: Erlbaum.
- BORNSTEIN, M. H., & LUDEMANN, P. L. (1989). Habituation at home. *Infant Behavior and Development*, 12, 525–529.
- BORNSTEIN, M. H., PÊCHEUX, M.-G., & LÉCUYER, R. (1988). Visual habituation in human infants: Development and rearing circumstances. *Psychological Research*, 50, 130–133.
- BOUCHARD, T. J., LYKKEN, D. T., MCGUE, M., SEGAL, N. L., & TELLEGEN, A. (1990). Sources of human psychological differences: The Minnesota study of twins reared apart. *Science*, 250, 223–228.
- CARON, A. J., & CARON, R. F. (1969). Degree of stimulus complexity and habituation of visual fixation in infants. *Psychonomic Science*, 14, 78–79.
- CARROLL, J. B. (1986). The measurement of intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. Cambridge, England: Cambridge University Press.
- COLOMBO, J., & MITCHELL, D. W. (1990). Individual differences in early visual attention: Fixation time and information processing. In J. Colombo & J. Fagen (Eds.), *Individual differences in infancy: Reliability, stability, prediction*. Hillsdale, NJ: Erlbaum.
- COOPER, L. A., & REGAN, D. T. (1982). Attention, perception, and intelligence. In R. J. Sternberg (Ed.), *Handbook of human intelligence*. Cambridge: Cambridge University Press.
- DEARY, I. J. (1992). *Auditory inspection time and intelligence*. Unpublished doctoral dissertation, University of Edinburgh.
- DIXON, R. A., & LERNER, R. M. (1992). A history of systems in developmental psychology. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook* (3rd ed.). Hillsdale, NJ: Erlbaum.
- FAGAN, J. F. (1992). Intelligence: A theoretical viewpoint. *Current Directions in Psychological Science*, 1, 82–86.
- FANTZ, R. L. (1964). Visual experience in infants: Decreased attention to familiar patterns relative to novel ones. *Science*, 146, 668–670.
- GARDNER, H. (1983). *Frames of mind*. New York: Basic Books.
- KOPP, C. B., & MCCALL, R. B. (1980). Stability and instability in mental test performance among normal, at-risk, and handicapped infants and children. In P. B. Baltes & O. B. Grim, Jr. (Eds.), *Life-span development and behavior* (Vol. 4). New York: Academic Press.
- KUHN, D. (1992). Cognitive development. In M. H. Bornstein & M. E. Lamb (Eds.), *Developmental psychology: An advanced textbook* (3rd ed.). Hillsdale, NJ: Erlbaum.
- MCCALL, R. B., & CARRIGER, M. S. (1993). A meta-analysis of infant habituation and recognition memory performance as predictors of later IQ. *Child Development*, 64, 57–79.
- PLOMIN, R. (1990). The role of inheritance in behavior. *Science*, 248, 183–188.
- SLATER, A. M. (1990). Infant development: The origins of competence. *The Psychologist*, 3, 109–113.
- SLATER, A. M., MORISON, V., & ROSE, D. (1983). Locus of habituation in the human newborn. *Perception*, 12, 593–598.
- WACHS, T. D. (1992). *The nature of nurture*. Newbury Park, CA: Sage.

MARC H. BORNSTEIN

INFANT TESTS AS MEASURES OF EARLY COMPETENCE

From the earliest stages of the scientific study of child development, there has been a strong interest in methods for charting the typical, or normal, course of early mental and behavioral development, beginning shortly after birth. For many child development specialists, the hope was that such normative data concerning developmental changes in the first two years of life could be utilized to measure individual differences in behavioral development, that is, to distinguish infants or toddlers who appear to be developing normally from those whose development seems to be either delayed or advanced. Some of these efforts were carried out by pediatricians such as Ar-

nold GESELL at Yale University, who was particularly interested in identifying delays in early mental and motor development so that appropriate remedial intervention might be provided. Other pioneering efforts were carried out by such child psychologists as Psyche Cattell, Nancy BAYLEY, and Ruth Griffiths. These efforts led to the development of a number of widely used infant tests or scales, intended to measure individual differences in mental and behavioral development during the first 24 to 30 months of life.

Infant tests have been employed by child psychologists and pediatricians for several different purposes. In clinical settings, they are used to evaluate the developmental status of individual infants or toddlers about whom there is concern regarding the possibility of abnormal or delayed development and also to evaluate changes in such children's developmental status after remedial treatment. Infant tests have also been widely used in research concerned with the effects of various risk conditions on early mental and behavioral development (e.g., malnutrition, low birth weight, exposure to drugs, unfavorable child care experiences, etc.) as well as in evaluating the results of wide-scale early intervention programs. They have also been employed in more theoretically oriented studies dealing with the continuity of early mental development and with the predictability of later intelligence from infant test scores in the first and second year of life. The latter issue is one about which there is still considerable controversy in the early 1990s.

This entry will first outline major features common to all infant tests and describe some of the most commonly used scales. Attention will then be directed to several general issues concerning the value of infant tests: (1) do they provide accurate or reliable measures of behavioral development during the first two years? (2) do they provide valid measures of early mental and behavioral development predictive of levels of intellectual functioning later in childhood? and (3) how useful are infant tests, given their advantages and disadvantages?

COMMON FEATURES OF INFANT TESTS

The various infant tests have several features in common in terms of their general content, their administration, and the nature of the scores obtained.

1. All infant tests are made up of a rather wide variety of items, specifying particular behaviors assumed to reflect important aspects of early development and typically observable at various age levels as children develop from birth through 24 to 30 months of age. These expectations of typical behaviors at different ages are based on extensive standardization data collected on representative samples of infants at various ages. For example, one would expect that an average 3-month-old would smile readily in response to a friendly examiner, an 8-month-old would be able to sit up without help, a 12-month-old would understand a few simple words like "Mommy" or "cookie," and a 19-month-old would place a round and square block in a form board.
2. Some of the behaviors constituting test items are elicited by the examiner following standard instructions and scored according to specific directions (e.g., "Put the block on the cup."). Other items involve behaviors that can be naturally observed during the course of the examination (e.g., infant spontaneously uses language or gesture to make wants known). In some instances, the mother's report may be used to score an item. Infants are usually tested with the mother or caregiver present, in an informal, relaxed setting. The examiner must be skillful in encouraging the baby to become engaged in the various test activities and must be aware of when the infant is too irritable, sleepy, or shy to allow a fair assessment.
3. When the examination is complete, the summary score basically reveals how the infant's responses compare with those of average children of that same age. The score, typically expressed as a developmental quotient (DQ), can be regarded essentially as a ratio of a child's performance or developmental age over the child's chronological age. Thus, if a 12-month-old child's performance is at the level typically shown by 15-month-olds, the DQ would be $15/12$ or 125; if performance was at the 9-month level, the DQ would be $9/12$ or 75, and if at the 12-month level, the DQ would be 100. As in the case of the intelligence quotient (IQ) employed with intelligence tests for older children, DQs between approximately 90 and 110 are considered to be within the average or normal range.

TYPICAL CONTENT OF INFANT TEST ITEMS

Most infant tests contain items representing the following broad behavioral domains:

1. *Gross motor development*: increasing control of posture and locomotion (e.g., holding head erect, sitting up without help, standing alone, walking without help, running, walking up and down stairs, jumping, etc.)
2. *Fine motor development*: use of fingers and hands for grasping and manipulating objects as required in everyday living (e.g., picking up small objects and placing them in containers, building a block "tower," use of spoon and cup in feeding self, placing cover on container, etc.)
3. *Perceptual-cognitive development*: attention to sights and sounds, ability to make relevant visual and auditory discriminations (e.g., discriminating between familiar and unfamiliar voices and faces, noticing when a change has been made in series of toys or pictures), ability to solve simple "problems" that require attention, memory, anticipation, and the beginnings of representational thinking (e.g., searching for an object that has been hidden, figuring out how to get a toy that is out of reach, remembering where a toy was hidden after some delay, placing blocks of different shapes in correct holes after form board has been reversed, etc.)
4. *Language comprehension and expression*: Language and intellectual development are closely related, so that many items included here also reflect the infant's emerging intellectual competencies. "Expressive" language items include cooing and babbling, using single words for objects or people, putting words together to form sentences, etc. "Receptive" items include responding appropriately to familiar words, carrying out simple requests, pointing to pictures or objects named by examiner, etc.
5. *Social/emotional development*: smiling at friendly examiner, showing pleasure in "peekaboo" or imitation games, expressing displeasure when toy is taken away, showing affection to familiar people, etc.

Most infant tests provide separate development quotient scores for different areas of development, as well as an overall DQ. The way specific items are clustered into the different domains and the way the

names are assigned to the various "scales" vary considerably, however, from one infant test to another.

DESCRIPTIONS OF COMMONLY USED INFANT TESTS

1. *Gesell Scales*: (4 weeks to 5 years) This is one of the earliest and initially most widely used infant assessment methods, particularly among pediatricians and neurologists concerned about general developmental delay. Developed at Yale University by Arnold Gesell beginning in the early 1930s, these scales provide scores in four areas: motor behavior (both gross and fine motor development); adaptive behavior (including manipulation and utilization of objects in a cognitively guided manner); language behavior, both "receptive" as well as "expressive"; and personal-social behavior (including appropriate social response to people, as well as self-feeding and bladder/bowel control). Although considered the "grandfather" of later infant tests, the Gesell Scales are not being as widely used in the late twentieth century since more recent tests provide better "norms" and are more advanced in terms of measurement technology.
2. *Cattell Infant Intelligence Scale*: (2 to 30 months) The Cattell test was developed by the psychologist Psyche Cattell at the Harvard University School of Public Health in 1940. It contains essentially no gross motor behavior content but does include some fine motor and object manipulation items, with the major emphasis being on perceptual-cognitive and language development. The test provides a single overall DQ or "IQ" score and continues to be used because of its briefer administration time and somewhat more focused concentration on early precursors of "mental" or "intellectual" development.
3. *Griffiths Mental Development Scale*: (1 month to 2 years) Developed by Ruth Griffiths in London in 1954, this is the only infant test standardized on British children. It yields both an overall DQ and separate scores for five subscales: locomotor, personal-social, hearing and speech (receptive and expressive language), hand and eye (fine motor and use of hands), and performance (object exploration, "problem solving"). Although not widely used in

the United States, the Griffiths Scale is frequently used in research in Europe, Africa, and Central and Latin America.

4. *Bayley Scales of Infant Development: (2 to 30 months)* The Bayley Scales have been the most widely used infant test in the United States. The test was developed by Nancy Bayley, a child psychologist working at the University of California at Berkeley and at the National Institutes of Health, beginning in the 1930s. Formal publication of the scales occurred in 1969, based on a standardization sample of some 1,200 children. As of the early 1990s, the Bayley Scales were being revised and restandardized by the Psychological Corporation. The Bayley provides two overall scores. The Mental Development Index (MDI) includes items reflecting perceptual-cognitive, language, and social development, as well as some involving object manipulation skills (as in the Cattell). The Psychomotor Development Index (PDI) is based on both gross and fine motor behavior, including hand use.
5. *Infant Mullen Scales of Early Learning: (2 to 36 months)* This 1991 infant test was developed by Eileen Mullen, a developmental psychologist and early childhood educator in Rhode Island, and standardized on a national U.S. sample of 1,231 children. The scales are intended to be particularly helpful in identifying delays that might warrant remedial intervention in five areas represented by the following subscales: gross-motor base, visual receptive/expressive organization (including fine motor and object manipulation items), and language receptive/expressive organization. It is also possible to combine the four visual and language organization scales into an overall composite Mental Development Index. As of the early 1990s, the Mullen Scales have not been sufficiently widely used to allow for extensive evaluation by critical reviewers.

ACCURACY OR RELIABILITY OF INFANT TEST MEASURES

Any infant test or procedure for measuring individual differences should be accurate or "reliable" in the sense of giving reasonably consistent results when the test is repeated after a short time interval. Infant tests have been shown to provide quite reliable measures of infant development, with test-retest correlations for

measures several weeks apart ranging from the .60s around 3 months of age to the .80s beginning around 12 to 15 months of age (Horner, 1980; McCall, 1979). Despite these satisfactory reliabilities, however, most specialists, when using infant tests for clinical evaluation of a particular infant whose development may be of concern, would want to administer the test more than once.

A related question has to do with the stability or consistency of infant test measures of development over longer time intervals during the first 24 months of life. Evidence from a number of studies indicates reasonable stability over periods of 3 to 6 months (r 's around .45 to .60), with correlations tending to be lower the earlier the testing and the longer the time intervals between tests (Honzik, 1983; McCall, 1979, 1983). There is some evidence to suggest that the stability of infant DQs may be considerably higher for some children than others, depending on constitutional factors, differential exposure to environmental changes or health risks, etc. (Wilson, 1983).

WHAT DO INFANT TESTS MEASURE?

Content Validity. The question of what is measured by infant tests can be addressed in part by consideration of the content of the scales. As is apparent from the previous description of the behavioral items and subscales typically incorporated in infant tests, the scores obtained reflect a number of what have long been considered significant aspects of early mental and behavioral development: gross and fine motor behavior, perceptual-cognitive and adaptive competencies, receptive and expressive language, and social behavior. It should be noted that the overall DQs (as well as some of the subscale scores) are typically based on a rather heterogeneous composite of different items, and thus provide *aggregate* measures of the infant's developmental status, rather than "pure" measures of specific intellectual, language, or social competencies. Moreover, because of the rapid pace of early development, the cluster of behavioral items on which the DQs are based tends to be quite different as one moves from the early months to the second year of life (Burns, Burns, & Kabacoff, 1992).

Concurrent Validity. Considerable empirical evidence that infant tests provide measures of significant dimensions of early mental and behavioral devel-

opment comes from numerous studies of the influence of various biological and socio-environmental risk conditions. For example, many investigations of the effects of low birth weight or prematurity utilizing infant tests have shown significant developmental delay in infants and toddlers, particularly if the degree of prematurity is substantial (e.g., Eckerman, Sturm, & Gross, 1985). (Prospects for subsequent "catch-up" development for such children are quite good in favorable home environments.) Infant tests have also proved very useful for evaluating the early developmental effects of malnutrition, prenatal drug exposure, congenital anomalies, and so on (Goodman, 1990; Honzik, 1983). Similarly, the socioeconomic differences in intellectual performance commonly observed in preschool children in many parts of the world begin to be discernible in infant test scores in the latter part of the second year of life (Golden & Birns, 1983). To the extent that infant tests have been able to detect hypothesized developmental effects of biological and socio-environmental risk conditions, as just outlined, they can be said to have shown evidence of "concurrent" validity.

Predictive Validity. One of the major concerns that has been raised about the value of infant tests is whether they are predictive of IQ levels later in childhood. The evidence here is quite clear, and the answers depend on one's objectives. If one wishes to predict later "intelligence" or IQ levels for children falling within the relatively "normal" range of IQ variation, infant tests generally tend to have little or no predictive value. The earlier the test, and the longer the interval preceding the follow-up IQ measure, the lower the correlations between them. By 18 to 24 months, however, one begins to find significant correlations (in the .50s) with 4-year IQ (Goodman, 1990; McCall, 1983; Siegel, 1989). This lack of predictive value within the normal range of intelligence should not be surprising given the previously mentioned heterogeneity of the aggregate infant test scores. Some evidence suggests that more focused measures of information processing, such as habituation or response to novelty, may represent more promising infant predictors of later intellectual functioning in the normal range (Bornstein, 1989; Goodman, 1990; Thompson, Fagan, & Fulker, 1991).

On the other hand, if the purpose is to identify infants and toddlers who appear to be developmentally

delayed, and who may have lower IQs later in childhood, then infant tests do have predictive value (Goodman, 1990; Largo et al., 1990). This conclusion is supported by two sorts of evidence. In correlational studies, when the samples employed include infants who were premature or developmentally delayed, thus increasing the range of IQ variation involved, significant correlations with later IQ are found even for infant tests given in the first year of life (Siegel, 1989). In addition, there is substantial evidence indicating that infants having DQs below 75 or 80 at 8 months are significantly more likely to have correspondingly low IQ scores at ages 4 to 8 years than infants whose DQs are above 80 (Holden, 1972; Ireton, Thwing, & Gravem, 1970; Siegel, 1989), particularly if the low DQ infants also come from more disadvantaged socioeconomic backgrounds (Siegel, 1989).

It may well be that the heterogeneity that makes infant tests poor predictors of later intelligence in the "normal" range actually contributes to their effectiveness as detectors of overall delay in behavioral development. When significant developmental delays appear early in life, they tend to be reflected not only in the perceptual-cognitive, language, and social domains but also in the gross motor and fine motor areas. The emergence of these motor competencies during early development represents a particularly significant aspect of development, since it tells us something about the maturation of important neurological functions during a period of very rapid growth and development. At the same time, infants with limited motor and manipulative competencies may have more limited experiences with the environment, thus restricting the range of available opportunities for learning. It should not be surprising, therefore, that there is suggestive evidence indicating that, in the first year of life, measures of motor, rather than "mental," development may be better predictors of later language delay (Siegel, 1989).

CONCLUSIONS

In summary, standardized infant tests provide reliable and useful measures of individual differences in overall behavioral development during the first two years of life. Because they provide aggregate or composite measures of development, they should not be

regarded as pure tests of early "intelligence" or of specific cognitive processes, capable of predicting later IQ scores in samples of normal children. They are, however, capable of identifying infants who may be significantly delayed intellectually later in childhood and of detecting the adverse effects of various socio-environmental and biological risk factors on early development. Thus, we can expect that infant tests will continue to be refined, with more emphasis on subscale scores, and that they will continue to be very useful for the purposes they were initially intended to serve.

(See also: INFANCY; INTERVENTION, INFANT AND PRE-SCHOOL.)

BIBLIOGRAPHY

- BORNSTEIN, M. H. (1989). Stability in early mental development: From attention and information processing in infancy to language and cognition in childhood. In M. H. Bornstein & N. A. Krasnegor (Eds.), *Stability and continuity in mental development* (pp. 147-170). Hillsdale, NJ: Erlbaum.
- BROOKS-GUNN, J., & WEINRAUB, M. (1983). Origins of infant intelligence testing. In M. Lewis (Ed.), *Origins of intelligence* (pp. 25-66). New York: Plenum.
- BURNS, W., BURNS, K. A., & KABACOFF, R. I. (1992). Item and factor analysis of the Bayley scales of infant development. In C. Rovee-Collier & L. P. Lipsitt (Eds.), *Advances in infancy research* (pp. 199-214). Norwood, NJ: Ablex.
- ECKERMAN, C. O., STURM, L. A., & GROSS, S. J. (1985). Different developmental courses for very-low-birthweight infants differing in early head growth. *Developmental Psychology*, 21, 813-827.
- GOLDEN, M., & BIRNS, B. (1983). Social class and infant intelligence. In M. Lewis (Ed.), *Origins of intelligence* (pp. 347-395). New York: Plenum.
- GOODMAN, J. F. (1990). Infant intelligence: Do we, can we, should we assess it? In C. R. Reynolds & R. W. K. Kamphaus (Eds.), *The handbook of psychological and educational assessment of children's intelligence and achievement* (pp. 183-208). New York: Guilford Press.
- HOLDEN, R. H. (1972). Prediction of mental retardation in infancy. *Mental Retardation*, 10, 28-30.
- HONZIK, M. P. (1983). Measuring mental abilities in infancy. In M. Lewis (Ed.), *Origins of intelligence* (pp. 67-105). New York: Plenum.
- HORNER, T. M. (1980). Test-retest and home-clinic characteristics of the Bayley scales of infant development in nine- and fifteen-month-old infants. *Child Development*, 51, 751-758.
- IRETON, H., THWING, E., & GRAVEM, H. (1970). Infant mental development and neurological status, family socio-economic status, and intelligence at age four. *Child Development*, 41, 937-945.
- LARGO, R. H., GRAF, S., KUNDU, S., HUNZIKER, U., & MOLINARI, L. (1990). Predicting developmental outcome at school age from infant tests of normal, at-risk, and retarded infants. *Developmental Medicine and Child Neurology*, 32, 30-45.
- MCCALL, R. B. (1979). The development of intellectual functioning in infancy and the prediction of later IQ. In J. Osofsky (Ed.), *Handbook of infant development* (pp. 707-741). New York: Wiley.
- MCCALL, R. B. (1983). A conceptual approach to early mental development. In M. Lewis (Ed.), *Origins of intelligence* (pp. 107-133). New York: Plenum.
- ROSS, G. (1989). Some thoughts on the value of infant tests for assessing and predicting mental ability. *Developmental and Behavioral Pediatrics*, 10, 44-47.
- SIEGEL, L. S. (1989). A reconceptualization of prediction from infant test scores. In M. C. Bornstein & N. A. Krasnegor (Eds.), *Stability and continuity in mental development* (pp. 89-103). Hillsdale, NJ: Erlbaum.
- THOMPSON, L. A., FAGAN, J. F., & FULKER, D. W. (1991). Longitudinal prediction of specific cognitive abilities from infant novelty preference. *Child Development*, 62, 530-538.
- WILSON, R. S. (1983). The Louisville twin study: Developmental synchronies in behavior. *Child Development*, 54, 298-316.

HENRY N. RICCIUTI

INFORMATION PROCESSING The information-processing approach to the study of human intelligence, which likens human to computer thought processes, represents a dramatic change from traditional approaches to studying intelligence. Where traditional approaches focus on identifying the broad dimensions of human abilities, information-processing

approaches focus on the thought processes underlying those abilities. Applications of information-processing methods promise to change how humans think about the very nature of intelligence. Where intelligence was once thought of as a broad, fixed mostly inherited trait, the information-processing approach is beginning to reshape the view of intelligence as a set of interacting processes operating on a variety of knowledge structures. The information-processing approach has been credited with being the key factor to revitalizing what had been a moribund field of scientific inquiry, one, in the words of Robert J. Sternberg, "as fashionable as a Nehru jacket." A new generation of intelligence researchers, schooled in the philosophy and methodology of information processing, has arrived, promising to change notions of human intelligence, how humans study it, how humans conceive of it.

THE INFORMATION-PROCESSING WORLDVIEW

The information-processing approach to the study of human intelligence is really two things, a virtual grab bag of techniques borrowed from experimental cognitive psychology (experimental procedures, mathematical modeling methods), and an attitude, a worldview, a way of thinking about how people think, learn, attend, solve problems, and made decisions. According to this worldview, one can understand thought by analyzing the flow of information a person processes, starting from an initial perception, through a series of cognitive processes, and resulting in either an observable response by the person or merely a change in the contents of the person's memory (i.e., what the person is thinking about). Consider the example of reading. The reader views a string of characters on the page (the perception); then mentally transforms the character string into a recognizable word (a cognitive process); retrieves the meaning of that word and, along with it, the associations that word has that have been earlier stored away (additional cognitive processes); and finally remembers the word in order to comprehend the sentence in which it is presented (another cognitive process resulting in a change in the contents of memory).

This all may seem rather obvious, but information-processing theories came about, in the 1960s and the 1970s, in reaction to "behaviorist" theories, inspired

by the psychologist B. F. Skinner, which view the human not as an active processor of information but rather as a passive, conditioned respondent to stimuli. Compared to behaviorist theories, information-processing theories are much more concerned with the ways in which incoming information is interpreted rather than simply with how people behave in response to information. Because of this emphasis on what is called cognitive mediation, information-processing theories are better suited than behaviorist theories to explaining complex human thought processes, such as those entering into performance of intelligence tests.

INFORMATION PROCESSING AS A SET OF METHODS

Just as important as the information-processing worldview in stimulating new ways of thinking about human intelligence were the techniques developed by information-processing psychologists working in the experimental tradition to study the mind (e.g., Neisser, 1967). The adaptation of these techniques by individual-differences researchers (e.g., Hunt, Frost, & Lunneborg, 1973; Sternberg, 1977) opened exciting new doors of inquiry and made possible the objective pursuit of issues only speculated on by previous generations of intelligence researchers. Before the 1970s, the predominant approach to studying intelligence was what has been labeled the "psychometric approach." This approach involves the administration of mental tests, the computation of a correlation matrix for the scores from those tests, the computation of an exploratory factor analysis of that matrix, and then the speculative psychological interpretation of the meaning of the derived factors. The adoption of information-processing methods allowed psychologists to move beyond simply interpreting broad factors and on to the study of the processes entering into the test-taking act itself.

THE COGNITIVE-CORRELATES METHOD

What exactly are these information-processing methods? Probably the most commonly employed technique for analyzing intelligence from an information-processing perspective is what J. W. Pellgrino and R. Glaser (1979)

called “the cognitive-correlates method,” which was pioneered by E. B. Hunt, N. Frost, and C. Lunneborg (1973). The method involves administering two sets of tests. One set consists of information-processing tasks, which, from previous experimentation, have been shown to reflect particular cognitive processes, such as memory retrieval, mental rotation, and scanning in short-term memory. The other set consists of one or more conventional intelligence tests that, from previous factor-analytic studies, have been shown to measure a psychometric factor, such as general ability, verbal ability, or spatial ability. By inspecting the pattern of correlations between the two sets of tasks, one is able to understand better the complex psychometric factor in more-basic information-processing terms.

The literature offers many examples of applying cognitive-correlates methodology to gain a clearer understanding of conventional abilities factors. Success stories can be found in the verbal realm, where researchers have demonstrated that the rather global construct of reading ability can be understood in information-processing terms as a combination of visual decoding ability along with retrieval speed from semantic long-term memory (Hunt, 1978; Jackson & McClelland, 1979). More recently, researchers have suggested that an additional important skill involved in reading is the ability to suppress irrelevant thoughts (Gernsbacher, 1990). Another, perhaps somewhat more modest success story exists in the learning realm, where general verbal learning ability has been shown to be largely accounted for by two independent factors, breadth of general knowledge and speed of searching long-term memory (Kyllonen, Tirre, & Christal, 1991). In the spatial realm, research has isolated particular processes, such as encoding, information-organization, and information-transformation processes as critical components of larger spatial-ability factors (Juhel, 1991). (See SPATIAL ABILITY.)

Since the time of the initial applications of cognitive-correlates methodology, there has been an expansion in how intelligence researchers have used the method. First, there has been an expansion in the concept of what an information-processing task is. At one time, the class of acceptable information-processing tasks was restricted to a fairly small set of cognitive tasks, such as letter matching and visual scanning, for which a large empirical data base existed and for which

detailed mathematical models of their processing requirements had been established. The picture has changed somewhat. Researchers now use a much wider variety of tasks and measurements as information-processing markers. In cognitive-correlates studies of GENERAL INTELLIGENCE, for example, researchers have used tasks such as pitch discrimination (Deary, Head, & Egan, 1989), line-length discrimination (Anderson, 1988; Kranzler & Jensen, 1989), and choice-reaction time (Jensen, 1982; Mathews & Dorn, 1989). There even have been studies that have used physiological measures, such as EEGs (Robinson, 1991) positron-emission tomography (PET) scans, and nerve-conduction velocity (Reed & Jensen, 1991) as information-processing measures.

In all cases, the key issue is whether the task (or measure) by aptitude correlation patterns will shed light on the aptitude being investigated. In studies of general ability, considerable interest has focused on trying to show that a fairly simple task or a fairly basic physiological measurement is correlated to the typically much more complex intelligence tests. The idea of general intelligence being reducible to something as simple as “speed of neural conductivity” has considerable intuitive appeal from a philosophy-of-science standpoint. After all, is not a driving goal in science to reduce a complex phenomenon to its simpler, more basic constituents?

A second expansion is that cognitive-correlates approaches are no longer being limited to analysis of conventional aptitude constructs. For example, D. Gitomer (1984) employed a cognitive-correlates design to determine the knowledge of skills that underlie proficiency in electronics troubleshooting. He identified a single critical underlying factor, proficiency in solving logic-gate problems. A follow-up to that study applied cognitive-correlates methodology to identify working-memory capacity as the single factor responsible for differences in the ability to learn how to solve logic-gate problems (Kyllonen & Stephens, 1990). Similarly, a cognitive-correlates approach has shown that working-memory capacity is the key factor underlying the ability to acquire skill in computer programming (Shute & Kyllonen, 1993). Finally, there has been expanded consideration of the cognitive-correlates method to more applied problems, such as aging (Rabbitt, 1988), learning disabilities (Swanson, 1990),

mental retardation (Vietze & Coates, 1986), and neuropsychological conditions (Robinson, 1991).

The expansion of the cognitive-correlates approach to include a wider variety of information-processing tasks and to attack a wider variety of complex intellectual skills and abilities greatly increases the power of the method as a tool for analyzing human intelligence. It is likely that work carried on within the cognitive-correlates tradition will continue to deliver insights into the nature of intelligence by successively relating it to more and more basic processing sources.

THE COGNITIVE-COMPONENTS METHOD

A second approach employed by information-processing researchers to study intelligence is what Pellegrino and Glaser (1979) have called the "cognitive-components" approach. Where the cognitive-correlates approach attempts to unpack intelligence indirectly, by examining its correlations with simpler, information-processing tasks, the cognitive-components approach unpacks intelligence directly, by isolating the processing stages that enter into test-taking performance. It does this by the use of a variety of techniques borrowed from experimental psychology, such as the "subtractive" and the "additive factors" methods.

The general approach was introduced to intelligence researchers by Sternberg (1977), who investigated analogical reasoning, which is what is called for in test items such as "Mother is to father as sister is to _____." Sternberg noted that scores from analogical reasoning tests are typically highly correlated with general intelligence. The question he was interested in pursuing was whether any of the information-processing stages associated with analogical reasoning could be shown to be the cause or the source of the correlation with intelligence. To tackle this problem, Sternberg broke performance on the analogical-reasoning test into simpler stages, such as "encoding" the stimulus terms (mother, father, etc.), "inferring" the relation between the first two stimulus terms, and "mapping" the relation found in the first pair (mother-father) onto the third term (sister) to identify the missing term (brother). Sternberg devised a series of partial tasks and employed a linear-modeling procedure to

identify the duration of each of the stages he hypothesized.

Having identified the duration of each of the stages (encoding, inferring, mapping, etc.), separately for each of the test takers, Sternberg was able to examine the reliabilities of the stages, their correlations with each other, and their correlations with external criteria, such as other measures of general intelligence or measures of overall task performance. In principle, the application of this kind of component breakdown should enable the investigator to make statements such as "The process that correlates most highly with intelligence is x ," where x is one of the stages. In practice, Sternberg's study consisted of too few subjects to allow for definitive statements on the importance of the components he had identified. Nevertheless, the impact of the study was substantial in getting intelligence researchers to think more expansively about how they could investigate the processes of intelligence.

An example of this kind of expansivity can be found in an analysis of what is considered one of the best measures of general intelligence, the RAVEN PROGRESSIVE MATRICES (RPM) test. P. A. Carpenter, M. A. Just, and P. Shell (1990) conducted a careful task analysis of what an information-processing system would have to do and would have to know in order to solve test problems. They formalized their thoughts in the form of a computer simulation that actually solved RPM problems. They then compared the computer's performance with the performance of human subjects to validate their computer simulation. They noted which problems humans had trouble with, and experimented with alterations of their computer simulation to find out what aspects of the simulation would have to be changed to mimic human performance. By doing this, they were able to determine that a key component of intelligent performance, as indicated by performance on the RPM, was the ability to keep track of goals and subgoals while solving problems.

Several issues are raised by the application of the cognitive-components approach. One is that it is important to have a good information-processing model of the task analyzed. Both Sternberg and Carpenter, Just, and Shell (1990), for example, cited a number of studies of reasoning (on their analogies and RPM-like tasks, respectively) from the cognitive psychological

literature and were able to develop information-processing model from those previous studies. For many of the key tests of intelligence, extensive research has been done from an information-processing perspective, and intelligence researchers can capitalize on this research for addressing issues pertaining to individual differences in the stages.

Another issue is that the investigator must be resourceful in developing a method for estimating stage performance or isolating processing success on a single stage of overall task performance. Carpenter, Just, and Shell relied heavily on verbal protocols, that is, examinees talking through the process of problem solving. Sternberg (1977) used the method of "precueing" to isolate stage performance. Precueing involves first presenting a part of the problem to be solved to the subject and then presenting the whole problem. If the whole problem involves two stages of processing and the precued portion only the first stage, then administering some problems with the precue and some problems without allows the investigator to determine the duration of the second stage.

A more common method for isolating processing stages involves varying features of the target aptitude test directly, that is, reconstructing the aptitude test so that it conforms to a standard experimental design. This is sometimes referred to as constructing a "faceted task." Consider a componential analysis of the mental-rotation task. Subjects try to determine whether two objects that appear in varying spatial orientations can be mentally rotated into congruence with one another (Shepard & Metzler, 1971). Researchers have studied the importance of the encoding and mental-rotation stages by varying the complexity of the encoding process (e.g., by using simple versus complex figures) and by varying the complexity of the rotation process (e.g., by comparing 0-, 30-, 60-, versus 90-degree-angularity disparity between the two figures). One can then observe how subjects respond to these changes in the items, and the pattern of responses can provide clues as to the importance of the different facets. For example, by carefully comparing subjects' responses over particular kinds of changes in difficulty, the investigator can calculate how quickly a subject can mentally rotate an object or how thoroughly a subject is able to encode an object. These parameters then can be compared to overall perfor-

mance in an attempt to determine the importance of the rotation stage or the encoding stage in driving overall performance (Pellegrino, Mumaw, & Shute, 1984).

THE COMPONENT-TRAINING METHOD

Besides the cognitive-correlates and cognitive-components methods, another method can be used to test whether a particular information-processing step is important for performance. The component-training method involves training subjects on a hypothesized component of an aptitude test. If such training increases performance on the aptitude test, that provides evidence for the importance of the hypothesized component. Sternberg (1984) has called this the "cognitive-training" method.

A rather clear example of this method can be found in a study by P. C. Kyllonen and associates (1991), who were interested in whether knowledge of an elaboration strategy on a paired-associates learning task was important for performance. They trained one group of subjects on the elaboration strategy, which resulted in about a 20 percent increase in performance over a no-training control group. This result suggests that knowledge of effective mnemonic strategies is an important component of associative learning. But an even more interesting finding was that the correlation between verbal ability and learning was higher for the group that was trained than it was for the untrained group. What this means is that some subjects come to the experiment with proper strategic knowledge and others do not, but that this knowledge is not related to their verbal ability per se; rather, it seems to be an almost idiosyncratic sort of knowledge. Thus, to get a clear idea about the relationship between verbal ability and learning, it is important to put everyone on a kind of common ground, to train everyone as to the importance of the tricks of how to do a paired-associates learning task.

The training method can be used in conjunction with other information-processing methods to identify critical components in intelligence performance. A nice example of this approach was a series of studies by J. R. Fredericksen, B. M. Warren, and A. S. Rosebery (1985). Initially, the investigators employed com-

ponential analysis and correlates methods to identify a set of specific information-processing components involved in reading ability. The components they identified were things like grapheme encoding, parsing sentence constituents, and retrieving and integrating word meanings. To validate the importance of these components in reading, Fredericksen, Warren, and Rosebery extensively trained a group of young readers on these specific components and found a resulting increase in their reading ability.

Component-training methods cannot be used as the final determinant of whether a particular component is involved in an intelligent performance. For one thing, subjects might become better at a particular component trained, but then have trouble transferring this knowledge or skill to the actual target performance. Research has shown, for example, that learners can receive extensive training on formulas such as $distance = rate \times time$ and still not be able effectively to use such knowledge to increase their problem-solving abilities. Also, it is important to keep in mind the causal direction of influence between components and the target performance. If one assumes that components determine performance in the target task, then changes in the components cause changes in the aptitude. But one might instead assume that components are simply indicators of the aptitude. In that case, changes in the aptitude cause changes in the components, but not necessarily the other way around. For example, if one trains people on how to take a particular test, that might not necessarily mean that one has changed the aptitude that the test is designed to measure.

THE COGNITIVE-TASK ANALYSIS METHOD

The researcher testing hypotheses about components of an aptitude construct must begin by developing a list of possible components. Such lists often originate from a kind of armchair analysis of what a person knows and does while solving an aptitude-test problem or from borrowing from the work of someone who has done such an analysis, but more-developed methods go beyond this kind of informal analysis. For example, there have been suggestions on how to get test takers to talk about the way they go about solving

problems both while they are solving them and after they are finished (Ericsson & Simon, 1980). A general problem is that people are not always very good at talking about what they can do. Some researchers have suggested getting two experts to solve a problem together, which has the effect of bringing out conflicting or alternative strategies, and generally to increase discussion (Hall, Gott, & Pokorney, 1993). Recording the eye movements made by subjects during problem solving sometimes is a useful way of thinking about what they are thinking about while they are engaged in the problem solving (Carpenter & Just, 1988). It is clear that the success of correlates, components, and training methods depends on the ideas about what components are involved in taking intelligence tests; cognitive-task analyses are designed precisely to suggest such components.

A CRITIQUE OF THE INFORMATION-PROCESSING APPROACH

Paradigm shifts in any field of scientific inquiry often lead to an initial flurry of activity denouncing the old ways, followed by a second wave of reconsideration, where the promises of the new approach are evaluated against the real progress the new approach has led to. It should come as no surprise that intelligence researchers are beginning to reconsider the true contributions of the information-processing paradigm shift. In a thoughtful critique of the information-processing approach to the analysis of intelligence, D. F. Lohman argued that the approach might have promised more than it ended up delivering. He noted that information-processing psychology has not "rescue[d] differential psychology from psychometrics, and return[ed] it to the mainstream of psychological research," as researchers at one time thought it would. This, he believes, was due to two false or misleading assumptions underlying the would-be marriage: Componential analysis would identify the true, psychologically correct sources of individual differences, and aptitude tests and cognitive tasks are interchangeable.

The problem with component scores (the product of componential analysis), according to Lohman, is that they have proven not to have construct or any other kind of validity. (A similar point was made by Brody, 1992.) Rather, the average or intercept score

from information-processing analyses has proven to be the most valid performance parameter. Because the average is what psychometrics already gave researchers and because what information processing promised to add was the component scores, the information-processing approach has not proved effective. The problem with treating aptitude tests as cognitive tasks, according to Lohman, is that aptitude tests do not allow for strategy differences, and finding strategy differences is one of the main potential contributions of the information-processing approach. In many ways, Lohman seems to have a good argument. It is easy to agree with him that components failed to be the true components of intelligence to the extent that many at one time thought they would.

Even so, it is important to acknowledge the contributions of the information-processing approach. For example, the oft-reported finding in the information-processing literature that component scores do not correlate very highly with intelligence-test scores is itself an interesting empirical finding, one that could have been anticipated. In other words, if component scores were what correlated with intelligence, then intelligence tests would have been more fragile than they have proven through the years to be. If the real essence of spatial ability were the rate of mental rotation, then the goodness of a test of spatial ability would be determined by how the test constructor sampled facet levels. The fact that one intuitively realizes that it does not matter how one samples (a test of rotation is a test of rotation, no matter how many degrees one's required to rotate) suggests that one knew the difference score was uninteresting all along.

Also, many in the field seem to believe that methods for identifying strategies are the most interesting contribution of information-processing psychology. Again, in retrospect, if this ever were the case, researchers have certainly been naive. The Raven progressive matrices certainly seem susceptible to alternative strategies. Yet, one should not have expected that someone could jump from average to superior intelligence, as represented by a Raven score, just by learning a better strategy. In the case of spatial ability, alternative strategies may exist, but in the end, how important is that? Are not pilots who can verbally encode their situation to a certain criterion level as well-off as those who spatially encode their situation to that

same criterion level? In the end, from a test-validity standpoint, is not how well (how accurate plus how much time taken) one gets the job done more important than how one goes about doing it?

It may prove in the end that the most important contribution of information processing is a set of psychological constructs (such as working-memory capacity) and a methodology for measuring those constructs. This is different from the view that information-processing psychology has suggested components and methods for isolating those components. It may be that there can be a synthesis of the sophisticated constructs of information-processing psychology and the methods for identifying and measuring those constructs taken from both information-processing and psychometric approaches. This would represent a true marriage between what L. J. CRONBACH called the "two fields" of psychology, the correlational and the experimental.

What about information-processing components? It is possible that these components still have a role to play, but one perhaps less grandiose than the one imagined at the outset. Specifically, components may serve as sources of difficulty on a particular intelligence test that can be systematically manipulated to make tests adaptively harder or easier. Several intelligence-testing programs are beginning to use components in exactly this fashion (Irvine, Dann, & Anderson, 1991). This is an intriguing turn of events, but one, in retrospect, that is perfectly sensible. The idea of components as originally outlined would have led to many more factors of intelligence than anyone has ever dared to suggest.

CONCLUSION

It may be that some of the early hopes of some information-processing researchers to banish the concept of intelligence, or at least to redefine it as a set of information-processing components, have not been realized. Conventional conceptions built over almost a century of research will not disappear overnight. Still, the information-processing revolution in intelligence testing has had, and will continue to have, a permanent impact on the way people view intelligence and the way they go about researching it. Information-processing psychologists have changed the conception of

what intelligence tests measure and have added many instruments to the arsenal for measuring it. These changes are leading to a greater unification of experimental and correlational traditions in psychology, and progress in the understanding of human intelligence will flow from developments in both fields.

BIBLIOGRAPHY

- ANDERSON, M. (1988). Inspection time, information processing and the development of intelligence. *British Journal of Developmental Psychology*, 6(1), 43–57.
- BRODY, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.
- CARPENTER, P., & JUST, M. A. (1988). The role of working memory in language comprehension. In D. Klahr & K. Kotovsky (Eds.), *Complex information processing: The impact of Herbert Simon*. Hillsdale, NJ: Erlbaum.
- CARPENTER, P. A., JUST, M. A., & SHELL, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review*, 97, 404–431.
- CHAIKEN, S. R., & YOUNG, R. (1993). Inspection time and intelligence: Attempts to eliminate the apparent movement strategy. *American Journal of Psychology*, 106, 191–210.
- DEARY, I. J., HEAD, B., & EGAN, V. (1989). Auditory inspection time, intelligence, and pitch discrimination. *Intelligence*, 13, 135–148.
- ERICSSON, K. A., & SIMON, H. (1980). Verbal reports as data. *Psychological Review*, 87, 215–251.
- FREDERICKSEN, J. R., WARREN, B. M., & ROSEBERY, A. S. (1985a). A componential approach to training reading skills, part 1: Perceptual units training. *Cognition and Instruction*, 2, 91–130.
- FREDERICKSEN, J. R., WARREN, B. M., & ROSEBERY, A. S. (1985b). A componential approach to training reading skills, part 2: Decoding and use of context. *Cognition and Instruction*, 2, 271–338.
- GERNSBACHER, M. A. (1990). *Language and comprehension as structure building*. Hillsdale, NJ: Erlbaum.
- GITOMER, D. (1984). *A cognitive analysis of a complex troubleshooting task*. Unpublished doctoral dissertation, University of Pittsburgh.
- HALL, E., GOTT, S., & POKORNEY, R. (1993). *PARI: A technique for interviewing experts*. Unpublished manuscript, Brooks Air Force Base, TX, Armstrong Laboratory, Human Resources Directorate.
- HUNT, E. (1978). Mechanics of verbal ability. *Psychological Review*, 85, 109–130.
- HUNT, E. B., FROST, N., & LUNNEBORG, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 7). New York: Academic Press.
- IRVINE, S., DANN, P. L., & ANDERSON, J. D. (1991). Towards a theory of algorithm-determined cognitive test construction. *British Journal of Psychology*, 81(2), 173–195.
- JACKSON, M. D., & MCCLELLAND, J. L. (1979). Processing determinants of reading speed. *Journal of Experimental Psychology: General*, 108, 151–181.
- JENSEN, A. R. (1982). Reaction time and psychometric *g*. In H. J. Eysenck (Ed.), *A model for intelligence*. New York: Springer-Verlag.
- JUHEL, J. (1991). Spatial abilities and individual differences in visual information processing. *Intelligence*, 15(1), 117–137.
- KAIL, R., & PELLEGRINO, J. W. (1985). *Human intelligence: Perspectives and prospects*. New York: W. H. Freeman.
- KRANZLER, J. H., & JENSEN, A. R. (1989). Inspection time and intelligence: A meta-analysis. *Intelligence*, 13, 329–348.
- KYLLONEN, P. C., & STEPHENS, D. L. (1990). Cognitive abilities as determinants of success in acquiring logic skill. *Learning and Individual Differences*, 2, 129–160.
- KYLLONEN, P. C., TIRRE, W. C., & CHRISTAL, R. E. (1991). Knowledge of processing speed as determinants of associative learning. *Journal of Experimental Psychology: General*, 120, 89–108. (Also [1989] Tech. Paper 87–68, Brooks Air Force Base, TX, Air Force Human Resources Laboratory.)
- LOHMAN, D. F. (1989). Human intelligence: An introduction to advances in theory and research. *Review of Educational Research*, 59(4), 333–373.
- MATHEWS, G., & DORN, L. (1989). IQ and choice reaction time: An information processing analysis. *Intelligence*, 13(4), 299–317.
- NEISSER, U. (1967). *Cognitive psychology*. New York: Appleton.
- PELLEGRINO, J. W., & GLASER, R. (1979). Cognitive correlates and components in the analysis of individual differences. In R. J. Sternberg & D. K. Detterman (Eds.), *Human intelligence: Perspectives on its theory and measurement* (pp. 61–88). Norwood, NJ: Ablex.

- PELLEGRINO, J. W., MUMAW, R. J., & SHUTE, V. J. (1984). Analyses of spatial aptitude and expertise. In S. Embretson (Ed.), *Test design: Contributions from psychology, education, and psychometrics* (pp. 45–76). New York: Academic Press.
- RABBITT, P. M. A. (1988). The faster the better? Some comments on the use of information processing rate as an index of change and individual differences in performance learning disabilities. In I. Hindmarch, B. Aufdembrinke, & H. Ott, (Eds.), *Psychopharmacology and reaction time*. Chichester, England: Wiley.
- REED, T., & JENSEN, A. R. (1991). Arm nerve conduction velocity (NCV), brain NCV, reaction time, and intelligence. *Intelligence*, 15(1), 33–47.
- ROBINSON, D. L. (1991). On the neurology of intelligence and intelligence factors. In H. A. H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement* (pp. 145–161). Hillsdale, NJ: Erlbaum.
- SALTHOUSE, T. A. (1992). The information-processing perspective on cognitive aging. In R. J. Sternberg & C. A. Berg (Eds.), *Intellectual development* (pp. 261–277). New York: Cambridge University Press.
- SHEPARD, R. N., & METZLER, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171, 701–703.
- SHUTE, V. J., & KYLLONEN, P. C. (1993). *Modeling individual differences in programming skill acquisition*. Unpublished manuscript. (Also [1990] Tech. Paper 90-76, Brooks Air Force Base, TX, Air Force Human Resources Laboratory.)
- SPEARMAN, C. (1927). *The abilities of man: Their nature and measurement*. New York: Macmillan.
- STERNBERG, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities*. Hillsdale, NJ: Erlbaum.
- STERNBERG, R. J. (1984). Testing and cognitive psychology. *American Psychologist*, 36, 1181–1189.
- SWANSON, H. L. (1990). Intelligence and learning disabilities: An introduction. In H. L. Swanson & B. K. Keogh (Eds.), *Learning disabilities: Theoretical and research issues* (pp. 23–29). Hillsdale, NJ: Erlbaum.
- VIETZE, P. M., & COATES, D. L. (1986). Using information processing strategies for early identification of mental retardation. *Topics in Early Childhood Special Education*, 6(3), 72–85.

PATRICK C. KYLLONEN

INSIGHT *Insight* is a commonly used word in everyday speech. People often talk about having insight into a problem or insight into their behavior. Moreover, insightful problem solving, as opposed to more routine forms of problem solving, is viewed as being responsible for many of the world's greatest contributions (Gruber, 1979). But what exactly is insight, and how is it related to other psychological constructs, such as intelligence?

A review of the literature reveals that insight has been difficult for psychologists to define. This difficulty arises partly because the term *insight* can be used in two different ways. One way refers to a product; the other refers to a process. Psychologists generally agree that insight as a product refers to an idea or solution that is nonobvious and of high quality. There is far less agreement, however, on how to define insight as a process. To understand thinking and problem solving, it is important to understand insight as both a product and a process. A review of how insight has been assessed in the past, followed by a discussion of the processes involved in insight and how they relate to other psychological constructs, will help illuminate this elusive construct.

CONVENTIONAL VIEWS OF INSIGHT

There are two conventional views of insight: the *special-process* view, and the *nothing-special* view. The special-process view is most often associated with the Gestalt psychologists, who believed that insight occurs when the problem solver suddenly sees a problem in a new way. This new perception, or *spontaneous restructuring*, is often thought to be accompanied by a novel solution, and a feeling of “Aha.”

According to the special-process view, the mental processes needed to solve insight problems are different from the more conventional abilities required to solve problems of the kinds found on tests of intelligence, creativity, and conventional problem-solving ability. Insight problems, typically, are novel problems that do not require much prior knowledge for their solution. For example, Burke and Maier (1965) used a “hat-rack” problem to study the relationship between insightful problem-solving ability and other abilities, such as creativity and intelligence.

The hat-rack problem requires participants to build a structure strong enough to support a heavy coat, using only two boards and a C-clamp. The opening of the clamp is wide enough so that the two boards can be inserted and held securely when the clamp is tightened. Participants are asked to build the hat-rack in the center of a small room. In order to solve this problem, subjects must have the insight that the ceiling of the room is relevant. The hat-rack is built by clamping the boards together and wedging them between the floor and ceiling. The handle of the clamp serves as a hook for the coat.

When performance on the hat-rack problem was correlated with subjects' Verbal and Mathematical scores on the Scholastic Aptitude Test, as well as with their scores on tests of creativity, personality, and interests, none of the correlations was significant. In other words, whether people had the insight needed to solve the hat-rack problem was unrelated to their scores on standardized ability tests. Burke and Maier concluded that the abilities needed to solve insight problems may be different from the abilities required to solve more routine types of problems.

The Gestaltists believed that people's inability to produce an insightful solution to a problem is often due to their fixation on past experience. For example, Duncker (1945) gave subjects three cardboard boxes, candles, matches, and thumb tacks. The task was to mount a candle vertically on a screen so that it could serve as a lamp. The solution is to melt wax onto the top of a box, stick the candle into the wax, and tack the box to the screen. Subjects who were given boxes filled with the tacks, matches, and candles had a harder time solving the problem than did subjects who received the same supplies outside of the boxes. According to Duncker, seeing a box serve the function of a container made it difficult for subjects then to view the box as a support. In other words, fixation keeps individuals from changing their problem-solving sets, even when old procedures are not relevant to the present situations.

More recent research in cognitive psychology supports the special-process view. High feelings of confidence that one is converging on the solution to an insight problem seem to be negatively predictive of correct solution to insight problems but positively pre-

dictive of correct solutions to more standard problems (Metcalfe & Weibe, 1987). In other words, subjects who feel they are gradually getting closer to solving insight problems tend to arrive at incorrect solutions, whereas subjects who feel they are far from solving the problems and then suddenly feel they know the answers tend to give correct solutions. Insight problems appear to be correctly solved by a subjectively catastrophic process rather than by accumulative processes, which fits the Gestalt notion that insight involves a sudden realization of a problem's solution.

Unfortunately, the special-process view has some problematic aspects. First, this view does not specify what insight is. Calling insight an "unconscious leap in thinking" or a "short-circuiting of normal reasoning" does not identify just what insight is or explain how it is distinct from other problem-solving processes. Second, the bulk of evidence in support of this view is anecdotal rather than experimental, and for each anecdote that supports the view, there is at least one anecdote to refute it (Perkins, 1981). Finally, the special-process view is probably not specified enough to permit direct empirical testing. As a result, it is not clear that the view is even falsifiable in its present form.

In contrast to the special-process view, the nothing-special view proposes that insight is merely an extension of ordinary thinking (Perkins, 1981). Insights, according to this view, are merely significant products of ordinary mental processes. This would mean that the "insight" problems used by the Gestaltists are not really insight problems at all. Instead, such problems are mostly alleged to measure the recognition of problem-specific prior knowledge. For example, Weisberg and Alba (1981) asked subjects to solve "classical" insight problems such as the "nine-dot" problem.

In the nine-dot problem, a subject is given a 3×3 array of nine equally spaced dots and is asked to connect the nine dots with four straight lines without lifting pencil from paper. What was unique about Weisberg and Alba's design was that subjects were given an insight that is needed to solve the problem: they were told that the problem could be solved only by drawing the lines beyond the boundaries formed by the dots. Subjects had difficulty solving the problem even after they were given this insight into how to

restructure it. However, subjects were much better at solving the problem after they had been trained on highly similar ones. The investigators interpreted the results as suggesting that relatively specific knowledge about a given problem, rather than insightful thinking, is the key to successful problem solving. They conclude that the terms *fixation* and *insight* do not belong in theories of problem solving.

This example illustrates how arguments for the nothing-special view are essentially arguments by default. After repeated failures to identify a construct empirically, one can easily be tempted to ascribe the failure to the nonexistence of the construct. However, recent research (Davidson & Sternberg, 1984) indicates that a main reason psychologists have had so much difficulty in isolating insight is that it involves not one process, but at least three related processes. What are the processes involved in insightful thinking?

THE PROCESSES OF INSIGHT

According to research conducted by Davidson and Sternberg (1984), insight comprises three fundamental mental processes. Insightful thinking occurs when these processes are performed in situations where the individual does not have a routine set of procedures for solving a problem. The three processes are *selective encoding*, *selective combination*, and *selective comparison*.

Selective Encoding. Insight occurs in encoding when a person sees in a stimulus, or set of stimuli, one or more things that previously have been nonobvious. Significant problems generally present an individual with a vast amount of information, only some of which is relevant to problem solution. An insight of selective encoding involves restructuring one's mental representation so that information that was originally viewed as being irrelevant is now seen as relevant for problem solution.

Ignaz Semmelweis's discovery of the importance of asepsis is a famous example of a selective encoding insight. While on the staff of a hospital in Vienna, Semmelweis noticed that more women on the poor ward were dying from infection during childbirth than were women on the rich ward. He encoded that doctors washed their hands less frequently while on the poor ward, and he realized the relevance that this had for spreading puerperal fever.

Selective Combination. Insight occurs in combination when one connects elements of a problem situation in a way that previously has been nonobvious to the individual. The elements are all there and often easy to see; the difficulty lies in finding a procedure to link them.

Darwin's formulation of the theory of evolution seems to have involved an insight of selective combination. He had all the facts for a long time; what he finally discovered was how to put the facts together to form a coherent theory.

Selective Comparison. Insight occurs in comparison when one discovers a nonobvious relationship between new information and information acquired in the past. It is here that analogies, metaphors, and models are used to solve problems. An insightful person suddenly realizes that new information is similar to old information in certain ways, and then uses this information better to understand the newly acquired information.

Archimedes' theory of "specific gravity" is a famous example of a selective comparison insight. While trying to determine whether silver had been put into King Hiero's gold crown, Archimedes stepped into a bath. He noticed that the amount of water that was displaced was equal to the volume of his body that was under water. By drawing an analogy between his bath and the problem with the crown, Archimedes suddenly knew how to determine the purity of the metal. He could compute the crown's volume by placing it in water and measuring the amount of displaced water. The crown could then be weighed against an equal volume of gold.

Not every instance of selective encoding, selective combination, or selective comparison is an insight. To be referred to as insightful, these processes must seem to occur abruptly when they do occur, and once they have occurred, must result in a change in the solver's mental representation of the problem.

THE RELATION BETWEEN INTELLIGENCE, IQ, AND INSIGHT

Intelligence, IQ, and insight are closely intertwined concepts. Intelligence has been defined in many ways, but for present purposes, it will be defined as the ability to comprehend information, solve problems, and

make decisions in a variety of situations. IQ is a measurement of a part of intelligence—in particular, that part which applies to the comprehension, knowledge retrieval, problem solving, and decision making required in academic kinds of tasks. Although some investigators have considered IQ to be synonymous with intelligence, this view is uncommon among recent theorists of intelligence (see definitions in Sternberg & Detterman, 1986), most of whom regard conventional intelligence tests as measuring abilities that span only a narrow range of the full set of abilities that constitute intelligence. Insight is also a part of intelligence, namely, that part which is applied to the solution of problems that are ill-structured and nonroutine. IQ and insight both involve the processes of selective encoding, selective combination, and selective comparison (Sternberg, 1985). In the case of IQ, these three processes are used in familiar ways; the problem solver knows the procedures that are applicable for solving the types of problems found on IQ tests, and, consequently, the selection of information is fairly routine. In the case of insight, the problem solver does not have a familiar representation and set of procedures that can be used on the problem. A new representation and new operators are constructed through the processes of selective encoding, selective combination, and selective comparison. Some individuals are better able to have insights than are others; this difference is related to differences in intelligence. In other words, highly intelligent individuals are more likely spontaneously to apply the three insight processes than are individuals with average or below-average intelligence.

In conclusion, insight involves a transition from not knowing the solution to knowing the solution. This transition is experienced by the problem solver as occurring suddenly, and it results in a change in the solver's mental representation of the problem. Highly intelligent individuals are more likely than those of average or below-average intelligence to engage in insightful thinking.

(See also: INTUITION.)

BIBLIOGRAPHY

BURKE, R. J., & MAIER, N. R. F. (1965). Attempts to predict success on an insight problem. *Psychological Reports*, *17*, 303–310.

- DAVIDSON, J. E. (1986). Insight and intellectual giftedness. In R. J. Sternberg & J. E. Davidson (Eds.), *Conceptions of giftedness*. New York: Cambridge University Press.
- DAVIDSON, J. E., & STERNBERG, R. J. (1984). The role of insight in intellectual giftedness. *Gifted Child Quarterly*, *28*, 58–64.
- DUNCKER, K. (1945). On problem solving. *Psychological Monographs*, *58*, 5 (Whole No. 270).
- GRUBER, H. E. (1979). On the relation between “Aha experiences” and the construction of ideas. *History of Science*, *19*, 41–59.
- METCALFE, J. (1986). Premonitions of insight predict impending error. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*, 623–634.
- METCALFE, J., & WEIBE, D. (1987). Intuition in insight and noninsight problem solving. *Memory & Cognition*, *15*, 238–246.
- PERKINS, D. (1981). *The mind's best work*. Cambridge, MA: Harvard University Press.
- STERNBERG, R. J. (1985). *Beyond IQ*. Cambridge, England: Cambridge University Press.
- STERNBERG, R. J., & DETTERMAN, D. K. (EDS.). (1986). *What is intelligence? Contemporary viewpoints on its nature and definition*. Norwood, NJ: Ablex.
- WEISBERG, R. W., & ALBA, J. W. (1981). An examination of the alleged role of “fixation” in the solution of “insight” problems. *Journal of Experimental Psychology: General*, *110*, 169–192.

JANET E. DAVIDSON

INTELLIGENCE QUOTIENT (IQ) The intelligence quotient (IQ) is the index of human intelligence as measured by a test score. It typically is one part of a two-part determination of an individual's intelligence, used in conjunction with an appraisal of adaptive behavior based on the individual's competence in everyday living, such as his or her school or work performance.

The first scale for measuring intelligence was based on a test given to subnormal children in France by Alfred Binet and Théophile Simon in 1905. The scale characterized a person's ranking among peers (the mental age)—a mental level that corresponded to the actual age (the chronological age) of normal children whose performance (number of test items answered correctly) had been equaled on that test.

In 1912, William Stern expanded the concept of mental age into a quotient when he conceived the idea of dividing the mental age (MA) indicated by the examinee's actual earned test score age by the chronological age (CA) and multiplying by 100: IQ equals MA divided by CA times 100. Thus, an 8-year-old who correctly answers the same number of test items as does the average 8-year-old earns an IQ of 100 (8 divided by 8 times 100); a 10-year-old who correctly answers the same number as does the average 8-year-old earns an IQ of 80 (8 divided by 10 times 100); and a 6-year-old who correctly answers the same number as does the average 8-year-old earns an IQ of 133 (8 divided by 6 times 100).

In 1916, Stern's method was retained by Lewis Terman of Stanford University in his development of the Stanford-Binet Intelligence Scale, the U.S. adaptation of the 1911 version of the Binet-Simon Scale. The MA was retained in computing an IQ until the 1937 revision; in the 1960, and 1972, and 1986 revisions, the Stanford-Binet substituted a variant of the deviation quotient (DQ), first used by David Wechsler in his 1939 Wechsler-Bellevue Scale.

BIBLIOGRAPHY

- KAUFMAN, ALAN, S. (1990). *Assessing adolescent and adult intelligence*. Boston: Allyn & Bacon.
- MATARAZZO, JOSEPH D. (1972). *Wechsler's measurement and appraisal of adult intelligence: Fifth and enlarged edition*. New York: Oxford University Press.

JOSEPH D. MATARAZZO

INTERACTIONIST VIEWS ON INTELLIGENCE In psychology, the term *interactionist* is used to refer to those theoretical views that focus on the relationships between the person and the environment. Interactionist views on intelligence entail a set of perspectives on cognitive functioning as an ongoing process. Rather than emphasize commonly understood states of being (e.g., "being intelligent") and/or an interindividual comparative version (e.g., "Jimmy is more intelligent than Johnny"), the interactionist viewpoint leads to clarification of the question: In what ways do persons' mental processes operate so as to allow an adequate adaptation of the person to the

environment? Likewise, persons reorganize their environments by active efforts, the results of which lead to further needs to adapt oneself to the changed environment. The person's psychological functions are considered as means to the end of constant adaptation (see Valsiner, 1984). Hence, for those investigators who adhere to the interactionist perspective, the term *intelligence* is merely a general common-language label, rather than a scientific term, to refer to the functioning and emergence of cognitive (mental) functions.

A related, more general view on intelligence can be found under the label *contextualism* (see Sternberg, 1984, 1985). However, the interactionist view only partially overlaps with the latter in its meaning. Not every contextualist viewpoint is interactionist, but all interactionist approaches are necessarily contextualist. The interactionist approaches are distinguishable by the presence of explicit interactive feedback loops between the person and the environment, together with the recognition of the irreversible nature of time.

In modern psychology, the popularity of the study of cognitive functions has led to the opening of new alleys of research into the mental and affective spheres of human beings (Gigerenzer, 1992). These new directions include the return to the ideas from mental psychology of the end of the nineteenth century (Danziger, 1990), the increasingly widespread use of computer terminology, and a focus on the social environment that is the context for the particular processes of thinking and feeling (in researchers' jargon, usually called cognition and affect, respectively). Cognitive psychology has benefited from the social fame of cognitive science over the recent decades.

NONDEVELOPMENTAL AND DEVELOPMENTAL PERSPECTIVES

The main task of cognitive research, the development of an adequate understanding of how minds function in the contexts of life experiences, has been approached from two distinctly different perspectives. First, the "nondevelopmental" perspective leads to varied conceptualizations of how the mind functions in its present state of organization, called the "steady state." From this point of view, the mind is an information-processing device that can interact with a variety of everyday life contexts, discover problems to be solved in those contexts, and work toward their

solution (Neisser, 1976). The structure of the mental processes involved in these encounters is not expected to change with the experience involved, yet the interactive feedback loop is in action (see Miller, Galanter, & Pribram, 1960: "Test-Operation-Test-Exit," or TOTE, model).

In contrast, the "developmental" perspective leads investigators to conceptualize the process of gainful change (qualitative transformation) in the organization of the structure of mental processes. In the course of interaction with the environment, the cognitive processes of a person can undergo complex reorganization, both progressive, in the sense that novel processes that afford tackling new problems emerge, and regressive, in that cognitive functions are lost when not necessary. The general law of development is that of differentiation and dedifferentiation of organizational forms, via articulation and hierarchical control mechanisms (Werner, 1957). When applied to intelligence, the developmental perspective would prescribe a careful study of the emergence of cognitive structures in a person's efforts to solve different problems. The main question asked is how new hierarchical structures develop in the process of activity.

HISTORY OF INTERACTIONIST PERSPECTIVES

Interactionist perspectives on cognitive development antedate psychology's invention of the term *intelligence*, as these perspectives emerged from the context of evolutionary thought applied to mental functions of different animal species in the late nineteenth century (Romanes, 1888). A major question for evolutionary thinkers since then has been whether evolutionarily lower levels of animal species can demonstrate mental functions comparable to those at higher levels (e.g., humans) and whether the mental functions demonstrable at the higher levels can be explained by concepts that pertain to lower levels (see an account of this history in Gottlieb, 1992). The concept of "consciousness" and its regulatory role in the organism-environment relationship (see Morgan, 1892) was at the heart of the issue of interactionist perspectives on mental functions. In philosophy, the issue that underlay the emphasis on treating mental functions in an interactionist framework was "free

will" (Meumann, 1908). In our contemporary psychology of cognitive processes, this theme continues to be discussed under the label *intentionality*.

Contributions by Baldwin. The central intellectual figure in interactionist approaches to intellectual functioning was the American psychologist James Mark Baldwin (see Baldwin, 1930). His interactionist stance was based on his concept of a "circular reaction," which envisions that in organism-environment interaction the responses to external stimulation are turned into stimuli for further reactions, with novelty constantly resulting from this cycle. Based on the notion of circular reaction is Baldwin's main concept for explaining development, that of "persistent imitation," wherein the organism's reconstruction of externally given models by way of experimenting with their properties results in the creation of moderately novel outcomes. Baldwin (1911) thus relates intelligence to the persistent nature of creative imitation:

Imitation to the intelligent and earnest imitator is never slavish, never mere repetition; it is, on the contrary, *a means for further ends*, a method of absorbing what is present in others and of making it over in forms peculiar to one's own temper and valuable to one's own genius [p. 22].

The question of intelligence is solved by Baldwin developmentally, by explaining it through creative imitation in the person-environment relationships.

Piaget's Interactionist Perspective. The work of Jean PIAGET in the area of cognitive development is widely known but poorly understood (with some exceptions, notably Chapman, 1988) in contemporary psychology. Piaget built his conceptual system on the developmental perspective of Baldwin and in parallel with gestalt psychological thinking (see Köhler, 1921). Piaget's focus on intelligence is embedded in his genetic epistemology (Piaget, 1970) of development through efforts to restore harmony in relations between person and the world. That harmony is constantly violated, and the person's cognitive functions are oriented toward establishing its reorganization by way of a process called "equilibration," which entails two interdependent processes, the assimilation of new information to the previous knowledge structure and the accommodation of that structure to incoming information (see Piaget, 1971). Piaget's views on cogni-

tive development have been developed further in a number of directions (Beilin & Pufall, 1992).

Contributions of the Hamburg School. In the history of human sciences, the contributions to interactionist perspectives on intelligence on behalf of the cohort of researchers at the University of Hamburg in the 1920s is noteworthy. The “Hamburg school,” led by the theoretical orientations of William Stern in psychology, Jakob von Uexküll in theoretical biology, and Ernst Cassirer in philosophy, included a number of relevant developmental psychologists, such as Heinz Werner (1940) and Marta Muchow (Wohlwill, 1985).

William Stern (who also happens to be the author of the term *intelligence quotient*, or *IQ*) made a profound contribution to focusing psychology on the study of person–environment relationships and of the heterogeneity of organization of the person (Stern, 1906). From the perspective of the interactionist personology of Stern (1938), intelligence is

the personal capacity to meet new demands by making appropriate use of thought as a means. . . . It is the task of intelligence (in contrast to memory) to meet *new* demands imposed by life, by making appropriate use of the means of thought at hand. Hence being able to think is in itself not intelligence; intelligence is the selective application of the means in the right place and in the right way [p. 309].

Sociosemiotic Developmental Views. The cultural-historical views on human development that were developed in late 1920s by Lev VYGOTSKY (see Van der Veer & Valsiner, 1991) and Alexander LURIA (see Luria, 1979) provided the basis for sociosemiotic developmental views of intelligence. Therein, intelligence is viewed as an outcome of development from “lower” (involuntary) to “higher” (voluntary) psychological functions, by way of the construction of novelty through dialectical synthesis. The Vygotsky-Luria semiotic tradition has been developed further by the Belgrade school of semiotically based developmental psychologists (Ivic, 1978) and by the analysis of cognitive processes in terms of “voices” (also based on the literary scholarship of Mikhail Bakhtin; Wertsch, 1991). The sociosemiotic tradition reconceptualizes issues of intelligence in terms of a “zone of proximal

development,” or the set (“zone”) of psychological functions that are currently in the process of emergence, in contrast to the functions that have already emerged by the given time (Valsiner & Van der Veer, 1993; Van der Veer & Valsiner, 1991). This has led to a number of novel perspectives on the teaching/learning process in socially guided human development (Ignjatovic-Savic et al., 1988; Newman, Griffin, & Cole, 1989).

Activity Theories. Parallels with the sociosemiotic views of the cultural-historical school are found in activity theories of intelligence (see Oppenheimer, 1991). The contemporary emphasis on the study of socially situated teaching/learning processes (Rogoff, 1990; Rogoff & Lave, 1984) has grown out from the activity-oriented viewpoints. It has redirected psychologists’ and educationalists’ research interests toward the reality of human development in everyday settings and offers productive alleys for educational intervention.

CONCLUSION

The interactionist perspectives on cognitive processes have concentrated on the explanation of mental functioning as a process of adaptation. These perspectives have borrowed from Alfred Binet’s largely forgotten theoretical emphasis on case-study analyses of mental adaptation (see Binet, 1903) and developed it further in multiple directions of thought, which have remained conceptually separate from the widespread proliferation of intelligence-testing practices. In their orientation, these perspectives are close to the contemporary emphasis on ecological approaches to human development (Bronfenbrenner, 1979, 1989), systems-theoretic analyses of human psychology (Ford & Lerner, 1992), and the coconstructivist perspectives on psychological functioning (Wozniak, 1986). Currently emerging traditions of cultural psychology (Boesch, 1991; Cole, 1985, 1990; Shweder, 1990) constitute a productive basis for future advancement of interactionist perspectives on cognitive processes and their development.

(See also: VYGOTSKIAN THEORIES OF INTELLIGENCE.)

BIBLIOGRAPHY

- BALDWIN, J. M. (1911). *The individual and society*. Boston: Badger.
- BALDWIN, J. M. (1930). James Mark Baldwin. In C. Murchison (Ed.), *A history of psychology in autobiography* (Vol. 1). New York: Russell & Russell.
- BINET, A. (1903). *L'étude expérimentale de l'intelligence*. Paris: Schleicher Frères.
- BEILIN, H., & PUFALL, P. B. (EDS.). (1992). *Piaget's theory: Prospects and possibilities*. Hillsdale, NJ: Erlbaum.
- BOESCH, E. E. (1991). *Symbolic action theory and cultural psychology*. New York: Springer.
- BRONFENBRENNER, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- BRONFENBRENNER, U. (1989). Ecological systems theory. *Annals of Child Development*, 6, 185–246.
- CHAPMAN, M. (1988). *Constructive evolution*. Cambridge: Cambridge University Press.
- COLE, M. (1985). Mind as a cultural achievement: Implications for IQ testing. In *Eighty-fourth yearbook of the National Society for the Study of Education*. Chicago: National Society for the Study of Education.
- COLE, M. (1990). Cultural psychology: A once and future discipline? In J. Berman (Ed.), *Nebraska symposium on motivation* (Vol. 37). Lincoln: University of Nebraska Press.
- DANZIGER, K. (1990). *Constructing the subject: Historical origins of psychological research*. Cambridge: Cambridge University Press.
- FORD, D. H., & LERNER, R. M. (1992). *Developmental systems theory*. Newbury Park, CA: Sage.
- GIGERENZER, G. (1992). Discovery in cognitive psychology: New tools inspire new theories. *Science in Context*, 5(2), 329–350.
- GOTTLIEB, G. (1992). *Individual development and evolution*. New York: Oxford University Press.
- IGNJATOVIC-SAVIC, N., KOVAC-CEROVIC, T., PLUT, D., & PE-SIKAN, A. (1988). Social interaction in early childhood and its developmental effects. In J. Valsiner (Ed.), *Child development within culturally structured environments* (Vol. 1). Norwood, NJ: Ablex.
- IVIC, I. (1978). *Covek kao animal symbolicum*. Belgrade: Nolit (in Serbo-Croatian).
- KÖHLER, W. (1921). *Intelligenzprüfungen am Menschenaffen*. Berlin: Springer.
- LURIA, A. (1979). *The making of mind*. Cambridge, MA: Harvard University Press.
- MEUMANN, E. (1908). *Intelligenz und Wille*. Leipzig: Quelle und Meyer.
- MILLER, G. A., GALANTER, E., & PRIBRAM, K. H. (1960). *Plans and the structure of behavior*. New York: Holt, Rinehart, & Winston.
- MORGAN, C. L. (1892). The law of psychogenesis. *Mind*, 1 (n.s.), 72–93.
- NEISSER, U. (1976). *Cognition and reality*. New York: W. H. Freeman.
- NEWMAN, D., GRIFFIN, P., & COLE, M. (1989). *The construction zone: Working for cognitive change in school*. Cambridge: Cambridge University Press.
- OPPENHEIMER, L. (1991). The concept of action: A historical perspective. In L. Oppenheimer & J. Valsiner (Eds.), *The origins of action* (pp. 1–35). New York: Springer.
- PIAGET, J. (1970). *Epistémologie des sciences de l'homme*. Paris: Gallimard.
- PIAGET, J. (1971). *Biology and knowledge: An essay on the relations between organic regulations and cognitive processes*. Chicago: University of Chicago Press.
- ROGOFF, B. (1990). *Apprenticeship in thinking*. New York: Oxford University Press.
- ROGOFF, B., & LAVE, J. (EDS.). (1984). *Everyday cognition*. Cambridge, MA: Harvard University Press.
- ROMANES, G. J. (1888). *Mental evolution in man: Origin of human faculty*. London: Kegan Paul, Trench.
- SHWEDER, R. (1990). Cultural psychology—what is it? In J. W. Stigler, R. A. Shweder, & G. Herdt (Eds.), *Cultural psychology*. Cambridge: Cambridge University Press.
- STERN, L. W. (1906). *Person und Sache: System der philosophischen Weltanschauung*. Leipzig: Barth.
- STERN, W. (1938). *General psychology from the personalistic standpoint*. New York: Macmillan.
- STERNBERG, R. J. (1984). A contextualist view of the nature of intelligence. In P. S. Fry (Ed.), *Changing conceptions of intelligence and intellectual functioning*. Amsterdam: North-Holland.
- STERNBERG, R. J. (1985). *Beyond IQ: A triarchic theory of human intelligence*. Cambridge: Cambridge University Press.
- VALSINER, J. (1984). Conceptualizing intelligence: From an internal static attribution to the study of the process structure of organism–environment relationships. *International Journal of Psychology*, 19, 363–389.
- VALSINER, J., & VAN DER VEER, R. (1993). The encoding of distance: The concept of the “zone of proximal development” and its interpretations. In R. R. Cocking & K.

- A. Renninger (Eds.), *The development and meaning of psychological distance*. Hillsdale, NJ: Erlbaum.
- VAN DER VEER, R., & VALSINER, J. (1991). *Understanding Vygotsky: A quest for synthesis*. Oxford: Basil Blackwell.
- WERNER, H. (1940). *Comparative psychology of mental development*. New York: Harper.
- WERNER, H. (1957). The concept of development from a comparative and organismic point of view. In D. B. Harris (Ed.), *The concept of development*. Minneapolis: University of Minnesota Press.
- WERTSCH, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA: Harvard University Press.
- WOHLWILL, J. F. (1985). Marta Muchow, 1892–1933: Her life, work, and contribution to developmental and ecological psychology. *Human Development*, 28, 198–224.
- WOZNIAK, R. H. (1986). Notes toward a co-constructive theory of the emotion–cognition relationship. In D. J. Bearison & H. Zimiles (Eds.), *Thought and emotion: Developmental perspectives*. Hillsdale, NJ: Erlbaum.

JAAN VALSINER

INTERFERENCE The concepts of interference and inhibition date back to the late nineteenth century, and they played an important role in theories of learning and memory. It was not until the late twentieth century, however, that these concepts began to make a contribution to theories of intelligence. The basic premise of modern-day interference theorists is that many, if not most, intellectual pursuits are interference-sensitive, inasmuch as they involve either irrelevant stimuli or are likely to trigger irrelevant or inappropriate thoughts (Table 1). As a consequence, the ability to resist interference, by inhibiting or suppressing information that is not relevant to the task at hand, is a critically important dimension of intelligence. This entry discusses specific phenomena that illustrate the wide generality of this emerging perspective on intelligence.

INTERFERENCE-SENSITIVE TASKS

Attention. One of the most important functions of the intellect is its role in selective attention. At any given moment our senses are exposed to a vast array

of stimuli, only a portion of which are relevant to the task at hand. To function effectively, therefore, it is necessary to ignore or suppress irrelevant, potentially interfering information.

Laboratory measures of selective attention index the ability to focus, divide, or maintain attention in the presence of irrelevant stimuli. In the Stroop Test, for example, the individual is required to name the color of the ink in which an incongruent word is written (e.g., the word *red* may be printed in green ink) as quickly as possible. The amount of time needed to do so compared with some other timed measure (e.g., naming colors on a color chart) provides an estimate of resistance to interference.

The results of studies using such tasks demonstrate both individual and developmental differences in selective attention. Stroop interference declines with age from 7 years to adulthood, remains reasonably stable over the young-to-middle-adult years, and increases for older adults (Davies, Jones, & Taylor, 1984). In general, these findings are consistent with everyday observations suggesting that both younger children and older adults are more easily distracted by external events than are younger adults. Further, individual differences in selective attention tasks are correlated with more general measures of intellectual ability. For example, subjects who were more resistant to interference in a selective attention task containing an irrelevant dimension also tended to score higher on the Scholastic Aptitude Test, a widely used college entrance test (Smith & Baron, 1981).

Comprehension. Individuals differ in many ways, including the ability to comprehend written and spoken language. Further, studies have shown that skill in comprehending written and spoken stories correlates highly with skill in comprehending nonverbal picture stories. Thus, there appears to be a general comprehension skill that extends beyond language and into many of our everyday activities (Gernsbacher, Varner, & Faust, 1990). In addition, performance on most aptitude and general ability tests, particularly paper and pencil tests, such as the verbal portion of the Scholastic Aptitude Test, are highly dependent on comprehension skill.

New evidence suggests that less skilled comprehenders, including young children and older adults, are

TABLE 1
Three sources of interference

<i>Type</i>	<i>Description</i>
Proactive	A previous event interferes with later thinking (A interferes with B).
Retroactive	New information interferes with a previous memory (B interferes with A).
Coactive	Two simultaneous events interfere with each other (A and B interfere with each other).

less able to resist the influence of irrelevant information than are skilled comprehenders. Both young children and older adults suffer from the tendency to carry over no-longer-relevant information into current processing. For example, older adults are more likely to maintain and consider interpretations of material they are reading than are younger adults, even when those interpretations have been contradicted or superseded by new information (Hasher & Zacks, 1988).

As to individual differences, less skilled readers exhibit greater interference from words that are irrelevant or inappropriate to the meaning of sentences as a whole than do more skilled readers. For instance, one second after reading a sentence such as “the man moved the piano,” less skilled fifth graders still show activation of semantically associated but contextually inappropriate words, such as “music”; in contrast, one second after reading the same sentence, more skilled fifth-grade readers demonstrate activation of only contextually relevant words, such as “heavy” (Merrill, Sperber, & McCauley, 1981). It does not appear, however, that less skilled readers are less cognizant of what is contextually appropriate than are more skilled readers. Instead, they appear to be less able to resist the influence of irrelevant associations. As a result, information that is inappropriate to the developing representation of the meaning of the text remains activated (Gernsbacher & Faust, 1991).

Unfortunately for those who are highly susceptible to text-based sources of interference, research suggests that many text passages may be troublesome. For example, a text that presents two or more related topics

in succession may produce abundant amounts of interference. Reading first about arteries and then about veins in a section on the cardiovascular system may seriously interfere with the reader’s understanding of the topic. The interference-prone reader is likely to confuse arteries and veins and be unable to respond correctly to questions about the properties of either (Dempster, 1988). Significantly, interference-prone college students tend to perform more poorly on a variety of measures of reading comprehension, such as the Scholastic Aptitude Reading Test, than do those who are less susceptible to interference (Dempster, 1985).

Problem Solving. It is becoming increasingly clear that errors in problem solving may arise because of misleading cues or irrelevant information. An individual may fail to reason correctly because the visual properties of the stimulus cause the person to “see” the problem incorrectly or because of some other irrelevant information.

One well-studied problem with misleading visual cues is conservation. Conservation refers to the ability to understand that certain properties of objects, such as mass, remain unchanged despite irrelevant changes in the appearance of one or more of the objects. Studies have found age-related improvements in conservation reasoning during childhood as well as declining performance with increasing age during adulthood (Dempster, 1992). Although these changes have been explained in different ways, the interference perspective suggests that age changes as well as individual differences in conservation performance reflect one’s ability to resist highly salient perceptual cues that con-

flict with the logically correct solution (Brainerd & Reyna, 1990; Dempster, 1992).

Errors in reasoning as a function of interference may also arise in problems that are strictly verbal, such as verbal analogies problems, which have played an important role in theories of intelligence. Consider the following:

Elephant is to small as _____ is to _____.

- (a) large:little (b) hippopotamus:mouse
(c)*lion:timid (d) turtle:slow

Although the problem has no obvious misleading cues, thinking-aloud protocols suggest that students who do poorly on this and related items often make mistakes that appear guided by irrelevant associations. Such students frequently pick (a) and, when asked to explain their choice, answer that "an elephant and small are opposites, and large and little are also opposites." Apparently, the inability to inhibit a superficially compelling association prevents them from developing the idea that an elephant is an animal and smallness is a quality that is not characteristic of that animal (Whimby, 1985).

A more direct link between intelligence and attention to irrelevant information when attempting to solve verbal analogies problems is suggested by a study that compared intellectually gifted and nongifted students. Gifted students allocated approximately the same amount of time to relevant novel information that preceded an analogies problem as did nongifted students, but quickly dismissed irrelevant novel information. Nongifted students, however, allocated as much time to irrelevant novel information as they did to relevant novel information (Marr & Sternberg, 1986).

CONCLUSIONS

Research in cognitive psychology suggests that resistance to interference, and by implication the capacity for inhibition, is a critically important dimension of intelligence. Research in a variety of domains, including selective attention, comprehension, and problem solving, provides compelling evidence that resistance to interference enters into a broad spectrum of intellectual phenomena. Further, the interference perspective represents a unifying framework for understanding diverse expressions of intellectual devel-

opment, cognitive aging, and individual differences. For example, it appears that increased resistance to interference during childhood is a major factor in the development of intellectual competence and that decreased resistance to interference in late adulthood is a major factor in declining mental ability later in life.

BIBLIOGRAPHY

- BJORK, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness* (pp. 309–330). Hillsdale, NJ: Erlbaum.
- BRAINERD, C. J., & REYNA, V. F. (1990). Inclusion illusions: Fuzzy-trace theory and perceptual salience effects in cognitive development. *Developmental Review, 10*, 365–403.
- DAVIES, D. R., JONES, D. M., & TAYLOR, A. (1984). Selective and sustained-attention tasks: Individual and group differences. In R. Parasuraman & D. R. Davies (Eds.), *Varieties of Attention* (pp. 395–447). New York: Academic Press. An exceptionally fine reference source on selective and sustained attention.
- DEMPSTER, F. N. (1985). Proactive interference in sentence recall: Topic similarity effects and individual differences. *Memory & Cognition, 13*, 81–89.
- DEMPSTER, F. N. (1988). Retroactive interference in the retention of prose: A reconsideration and new evidence. *Applied Cognitive Psychology, 2*, 97–113.
- DEMPSTER, F. N. (1991). Inhibition processes: A neglected dimension of intelligence. *Intelligence, 15*, 157–173. This article, which contains numerous references, is the first to deal directly with the role of interference in intelligence.
- DEMPSTER, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review, 12*, 45–75. A highly informative review of the interference perspective as it applies to intellectual development and aging.
- GERNSBACHER, M. A., & FAUST, M. E. (1991). The mechanism of suppression: A component of general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 245–262. This authoritative article makes a compelling case for the role of suppression (i.e., inhibition) in comprehension.
- GERNSBACHER, M. A., VARNER, K. R., & FAUST, M. E. (1990).

Investigating differences in general comprehension skill. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 16, 430–445.

- HASHER, L., & ZACKS, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 193–224). New York: Academic Press. A highly influential discussion of the inhibitory decline hypothesis in normal aging.
- MARR, D. B., & STERNBERG, R. J. (1986). Analogical reasoning with novel concepts: Differential attention of intellectually gifted and nongifted children to relevant and irrelevant novel stimuli. *Cognitive Development*, 1, 53–72.
- MERRILL, E. C., SPERBER, R. D., & MCCAULEY, C. (1981). Differences in semantic encoding as a function of reading comprehension skill. *Memory and Cognition*, 9, 618–624.
- SMITH, J. D., & BARON, J. (1981). Individual differences in the classification of stimuli by dimensions. *Journal of Experimental Psychology: Human Perception and Performance*, 1, 1132–1145.
- WHIMBY, A. (1985). Focusing on trainable *g*. *The Behavioral and Brain Sciences*, 8, 245–246.

FRANK N. DEMPSTER

INTERVENTION, INFANT AND PRESCHOOL Impelled by reports of failing test scores and poor academic performances by children from poor or low-literacy families, educators in the early 1960s implemented intervention programs that used a variety of models designed to increase the learning skills and ultimate academic achievements of children at risk for later school failure. Intervention models differed in the locale in which they were carried out (e.g., center versus home) and in the target person for the intervention effort. Some provided group instruction or learning opportunities for children. Others tried to empower families to become teachers of their preschool youngsters. A few of the models addressed emotional as well as learning needs of the child under the hypothesis that attention to cognitive curriculum alone would not provide the emotional basis for intellectual gains. Time of entry and duration of program varied. Some programs focused on infants, and program enrichment continued for two or more years.

Others targeted children in the year before they entered elementary school to provide an intellectual boost that would allow them to do well in kindergarten. Almost all the intervention and prevention programs were able to demonstrate short-term academic gains.

MODELS

Group Programs. The Syracuse Children's Center (Caldwell et al., 1970) was a pioneer infant intervention model designed to prevent the downward drift in IQ so typically found within the first several years among children growing up in disadvantaged environments. Sensitive to infant needs to build a secure emotional relationship with the mother, the Children's Center admitted babies to group care only after the first half-year, and enrichment was provided half-days for the first year of program. So cautious were the early infant interventionists that Mary Elizabeth Keister (1970) reported with relief that participation in her enriched university-based program did not harm infants intellectually.

Among the more narrowly conceived group programs to boost intellectual and language achievement was C. Bereiter and S. Engelmann's (1966) "pattern drill" model. Each teacher worked with a small group of disadvantaged 4-year-olds whose language scores fell below national norms. Children responded to specific questions with specific phrases that indicated concepts such as classification names or causal relationships. (Response examples are "If the block is big, it must be blue" and "No, this is not a book. This is a pencil.") Although early results suggested the efficacy of this model, later reports suggested that other, more child-development based models, had longer-term effects in sustaining cognitive gains. However, Gerstein & George (1991) reported that disadvantaged 5-year-olds who had participated in this direct-instruction model in kindergarten (reading, math, and language sessions in groups of six to ten preschoolers with a teacher) outperformed comparison students academically six years after participating.

The Frank Porter Graham Center in Chapel Hill, North Carolina, provided a high-quality, initially purely cognitively oriented group program from three months to school age, for African-American infants of

low-education single mothers. Some of the babies were fetally malnourished, and others were not. Program effects after three years were remarkable (Ramey & Gowan, 1990). Experimental infants who were normal at birth achieved Stanford-Binet IQ scores of 98.1 and fetally malnourished infants also randomly assigned to the educational program attained 96.4. In contrast, the group of normal infants randomly assigned to the control group and those control infants who were initially fetally malnourished had IQs, respectively, of 84.7 and 70.6. To attribute the cognitive gains to program educational efforts entirely is difficult because control mothers gradually showed less interest in their infants than experimental group mothers. Educational enrichment produces *transactional* effects, so that the parents of children whose learning has been given a boost may find the children more interesting and more receptive to learning interchanges at home. Thus, early educational intervention seems to have indirect as well as direct effects on children's early cognitive development. The goals of the ABECEDARIAN PROJECT were later modified to include more prosocial emphases when the program graduates were found to be fifteen times as aggressive as control children in kindergarten classrooms and settings, and aggression declined sharply in later program cohorts. In a further intervention to enhance the parent-as-teacher role, "contingent verbal responsiveness was an important component of the social mediated experiences that seem to be at the nucleus of influences on early cognitive development" (MacPhee, Ramey, & Yeates, 1984, p. 361).

The Perry Preschool Program (Berrueta-Clement et al., 1984) randomly assigned low-income African-American disadvantaged children with IQs between 60 and 90 either to program or to control groups. Attrition in the follow-up study at age nineteen was remarkably low. High-school graduation rates and scores on an adult test of functional competence were significantly higher for the Perry preschool group than for controls. Fewer preschool students were labeled *mentally retarded* than were control students, and program graduates spent slightly over half the number of years (of nonprogram children) in special-education classes. Although some program graduates did poorly academically, in general this program for poor children at cog-

nitive risk shows the long-term positive effects of two years (WAVE 1 children had only one year) of developmentally appropriate programming. By age twenty-seven, 71 percent of the program group had obtained a high school graduate degree or equivalent, compared with 54 percent of the controls, and the program group scored significantly higher on educational performance (Schweinhart, Barnes, & Weikart, 1993).

Tutorial Models. In most tutorial intervention models, a tutor came to the home to work directly with the infant to promote cognitive advancement. Infants typically gained about ten IQ points in comparison with controls after one year of in-home cognitive tutoring. In Palmer's model (Day & Parker, 1977), tutors worked with African-American male toddlers to boost concept development and knowledge of polar opposites, such as heavy-light, wet-dry, and over-under. One group of toddlers received carefully sequenced activities to promote concept learning through direct tutoring, and a second "discovery" group explored the materials in the presence of a responsive tutor who did not initiate teaching. After one year, the concept group performed better than the discovery group; but after several years, both of the program groups were indistinguishable from each other and superior in concept knowledge to control youngsters who had not experienced either curriculum.

Schaefer trained in-home tutors to present cognitively enriching learning games to poor African-American toddlers for a year (see Day & Parker, 1977, for elaborate descriptions of such early programs). Although their IQ scores were significantly raised in comparison with control children, the gains dissipated. By 6 years, there were no longer significant differences between program children and controls who had already had a year of elementary school. In that tutorial program, mothers were not required to learn any tasks to carry out with the babies nor were they encouraged to gain insight into the techniques and principles that the interveners were using.

Some tutorial models focused on teaching the mother how to enhance her infant's cognitive development.

Levenstein's (1988) Mother-Child Home Program (MCHP) targeted low-income housing-project families in Long Island for either one or two years of interven-

tion. Mothers were given a toy or book during each home visit and were shown how to use the materials to encourage language interactions with their toddlers. In both short-term and long-range outcome studies, IQ scores for all six cohorts of MCHP graduates were significantly superior to those of untreated or placebo-treated control groups and to their own pretest IQ scores. The average gain after two years was 17 IQ points. Postprogram score advantages lasted into fifth and eighth grades. Fifty-two younger siblings of MCHP children entered with IQ scores one-half standard deviation higher than the pretest scores of their older siblings. This fact suggests that mothers fostered the younger children's intellectual development through parenting skills learned in the program. This model has been replicated at many sites and has modest costs compared with programs that require preschool sites. The model promotes cognitive gains primarily for high-risk mothers labeled *hesitators* (those who do not have a high school diploma), rather than *strivers*, who have worked to attain the high school diploma. D. Olds has similarly noted that significant cognitive achievements were found in his home visitation program only for children of young, single, and poor mothers—those who were most at risk (Olds et al., 1985).

In Vermont, the Mother-Infant Transaction Program taught mothers of low birth-weight infants (LBW) to respond appropriately to infant cues. Despite biomedical vulnerability, "the intervention prevented cognitive lags among LBW children" (Achenbach, Phares, & Howell, 1990, p. 1672). The seven-year follow-up revealed that once social class status was controlled for, children of these trained LBW mothers achieved significantly higher cognitive scores on the KAUFMAN ASSESSMENT BATTERY FOR CHILDREN. The HIPPIY model (National Council of Jewish Women, 1982) was designed to combat continued educational disadvantage in Israel among preschoolers from North African immigrant families in comparison with preschoolers in Western families. This disadvantage persisted despite excellent preschool education available to all. Mothers received weekly in-home programmed instruction from home visitors who stress language, math, sensory, and perceptive skills presented in gamelike activities. By participating in the

activities, children developed habits of paying attention, concentrating, anticipating, and finding out what is expected of them in a learning situation. In the initial HIPPIY research, at the end of kindergarten, program children were doing better on the Boehm Test of Basic Concepts, which correlates highly with reading readiness. At the end of first grade, the HIPPIY home tutorial group was superior to control students in math. By the end of second grade, significant differences between home tutored children and control youngsters were evident in reading and math. By age 17 to 18 years, 69 percent of HIPPIY graduates were at correct grade levels compared with 53 percent of the others. This model is now in use in several cities in the United States, including Miami, St. Louis, and Little Rock.

Mixed Models. Variations in models include mixing tutorial and group learning situations and involving parents as teaching aides in classrooms (Honig, 1979). The Milwaukee Project recruited babies at birth from families in which low-IQ mothers lived in dilapidated housing in poor neighborhoods. Mothers learned literacy skills and job skills as laundry or dietetic aides, but they were not taught child-development knowledge or skills. One-on-one tutorial sessions were held with the infant in the home during the first months until the mother allowed the tutor to bring the infant to the learning center. Caregiver/infant ratios were 1 : 1 until a year; 1 : 2 until 15 months, and 1 : 3 by 18 months. Careful curricular planning of the educational component resulted in mean preschool IQs of 121.6 for experimental children compared to control IQs of 95.7. Children in this program scored thirty-nine IQ points higher than their older siblings had scored at the same age. Longitudinal results, however, were discouraging (Garber, 1988). By 120 months, experimental mean IQs were 104.2 and controls tested at 86.3. Although the experimental children retained an IQ advantage, by the end of high school they were showing the same pattern of academic failure, school drop out, and conduct problems as their peers in the inner-city public schools they attended. Three factors seem to account for the dissipation of intellectual advantage among program graduates: lack of educational continuity of special enrichment; lack of emphasis on empowering parents as

primary educators of their own children; and a public school milieu characterized by negative attitudes toward learning.

In the Houston PCDC (Parent-Child Development Center) demonstration program, Mexican-American infants were taught in a home-visitation program for an initial trust-building year, after which the children entered a group program to build language and cognitive skills. Cultural sensitivity in working with some families may mean "piggybacking" one model in sequence with another to advance child cognition.

A. McQueen and V. Washington (1990) randomly assigned low-income African-American mother-child pairs to one of three groups. All the children were already in preschool. Mothers were either in a control group, in a group in which mothers read a story a week to their child at home, or in a Parent Education Program (PEP) in which mothers learned how to teach their children, constructed educational toys and materials to stimulate verbal conversations, read stories and discussed them with the child, and had the child tell the story to classmates. Mothers in the PEP group also served as teacher assistants in their own child's classrooms. Only in the PEP group were posttest Peabody Picture Vocabulary Test mean scores higher than pretest scores. The PEP children's mean verbal PPVT intelligence scores reflected a gain of twenty months during a period of three months in contrast to the other two groups. Thus, simply attending a preschool program does not necessarily boost the cognitive scores of disadvantaged preschoolers. Gains are found, however, when parents are actively involved with specific language-focused instructions in working with their own children at home and in the classroom.

Omnibus Models. Some intervention programs were broadly based to ensure that cognitive gains would be more likely sustained after the program ended. The Syracuse Family Development Research Program (FDRP) was a long-term omnibus model whose goal was to sustain family functioning through home visitation and high-quality child care. Low-income, teenaged single mothers who gave birth before completing high school learned nutrition, life coping skills, story-reading, and Piagetian cognitive games to play with infants and preschoolers. From 6 months onward, the infants entered the Syracuse Children's Center group care program. The curriculum combined

ideas from Erik Erikson, Jean PIAGET, and language development theorists. For five years, children attended a mixed-age, open-education program with four major learning areas from which they could choose. Teachers also targeted specific skills in tutorial small groups available daily. Weekly home visits and monthly parent meetings provided active supports for parents. By 5 years, significantly more of the program children attained Stanford-Binet IQ scores above 89, when compared to their controls. Longitudinal follow-up ten years after graduation (Lally, Mangione, & Honig, 1988) revealed that FDRP graduates were significantly more likely to expect to be in school five years later. Females, but not males, were doing significantly better academically than control youth. A 6 percent juvenile delinquency rate for program graduates was significantly lower than the 22 percent rate for control youth. Problems for disadvantaged males in families with no strong male role models for intellectual and educational achievement were highlighted by the lack of academic superiority of FDRP male graduates in comparison with controls.

Even Start is a family-focused model that integrates early childhood education and adult literacy education for parents into a unified program that also provides a range of support services such as mental health referrals. The goal is to engender attitude and skill changes in parents in order to motivate long-term family support for children's learning (Administration for Children, Youth and Families, 1992, pp. 386-390).

In Project CARE, IQ scores from the Stanford-Binet tests were not significantly higher at 48 months for those children who had received educational care in the Abecedarian Project combined with a family education program for their high-risk families in comparison with children who had received either home visitation exclusively or had attended no program (Wasik et al., 1990).

The Consortium for Longitudinal Studies (1983) carried out systematic inquiry into the pooled effects through twelfth grade of a variety of model enrichment programs (tutorial and group) that targeted poverty or minority culture preschoolers. Across projects, program/control IQ score differences were 7.42 after program, 4.32 after one year, and 4.63 after two years of the program. The early education programs "produced an increase in children's IQ scores that lasted

for several years after the program . . . [but] the effect was not permanent” (p. 428). Program children did complete more years of schooling and were significantly less likely to be retained in the same grade than were control youngsters. Across the eight projects, the average rate of grade retention was 19.8 percent for program graduates and 32 percent for control students. In addition, “control subjects averaged 3.08 years in major academic [remediation] placements compared to 1.23 years for program participants” (p. 436). Finally, the preschool programs increased the rate of high-school graduation or equivalency diplomas in comparison with control children (Lazar et al., 1982). Despite significant differences in many of the models, all the curricula were successful in reducing school failure. As the consortium concluded, “It appears that a variety of curricula are equally effective in preparing children for school and that any of the tested curricula is better than no preschool program at all” (p. 442).

INTERVENTION PARAMETERS

Quality and Intensity of Intervention. Although intervention models differ in developmental rather than academic focus, in age of preschoolers served, and in length of program time, two qualities have tended to characterize models that have significantly raised preschool IQ. In an analysis of seventeen early intervention programs that involved random assignment of program and control preschoolers, the best predictors of IQ gains of enrolled preschoolers were quality of program, intensity of teacher-child contact, and breadth of program (Bryant & Ramey, 1987). In the National Day Care Study in Atlanta, Georgia, children were randomly assigned to classrooms varying systematically in quality factors. The growth of their cognitive scores was most closely tied to the quality of the child care centers they attended (Ruopp et al., 1979). Analyzing the cognitive and language data in a study of child care programs in Bermuda, K. McCartney (1984) similarly concluded that high-quality programs appeared to have a positive effect on language development. When caregivers showed more individual attention to children, engaged them in more structured activities, and interacted with them in smaller groups, the Preschool Inventory (a

measure of school readiness) and PPVT scores were higher. Cognitive development was higher in centers where this was a specific staff goal and the focus was more on individual development than on group experience. Such findings make necessary a cautious response to conflicting data about effects of intervention programs. Intellectual gains can be optimally predicted when programs meet specific criteria of individualized cognitive attention in interactions with preschoolers and when developmentally appropriate activities are carefully structured for the children.

Programs that emphasize one particular aspect of cognitive content will likely result in children making significant advances in that area. In a study that emphasized visual-spatial problem-solving skill, program children showed superior performance in this area on the Griffiths IQ test (Fowler & Khan, 1974).

Research Design Difficulties. Several factors make it difficult to assess the long-term sustained effects of enrichment programs designed to have an impact on intellectual development of preschool children at risk for school failure (Clarke-Stewart & Fein, 1983). Different programs use different measures such as rate of dropout from school, specific cognitive or IQ or language tests, and rate of assignment to special education.

Social class impacts on extent of intellectual gains. Fowler and Khan (1974) reported 19 and 16 IQ point gains respectively, for middle and lower socioeconomic (SES) preschoolers. After two years in the Syracuse Children’s Center, middle and low SES infants had gained 17 and 9 IQ points, respectively (Caldwell et al., 1970). Of special interest is the finding that by 30 months, low SES infants showed a rise in IQ (above 105), and matched, exclusively home-reared controls showed the IQ decrease typical of low SES infants.

A troubling finding is that often disadvantaged preschoolers still experience school difficulties despite IQ gains. For example, migrant workers’ children in Woolman’s microsocial learning environment program (see Consortium for Longitudinal Studies, 1983) within a year improved their Wechsler IQ scores from 79.2 to 88.8. Although the difference is significant, the latter IQ score does not bode well for later school success.

Subject attrition over time impacts on longitudinal follow-up of children. Control families who are less

attached to program goals and personnel may be most difficult to locate and only the most highly successful of control children may be available after several years. Sample attrition is often most severe in the most stressed families. Thus, findings of superior cognitive functioning among program children in longitudinal follow-ups make the data difficult to interpret. The children who may not benefit are least likely to be in the follow-up sample. In other research designs, more needy children are enrolled in enrichment groups, despite the fact that families are carefully matched at the outset. In these cases also, comparisons will be difficult because the program families and children may have had more life stresses and difficulties than those control families motivated enough to be available for project assessments over the years.

Enrichment programs may lack funds for a control group either by random assignment or carefully matched controls. For example, in family focused enrichment programs chosen for their excellence, measures of intellectual achievement were primarily collected on program children only. B. Goodson, J. Swartz, and M. Millsap (1991) noted pre-post cognitive gains in reviewing seventeen promising programs for disadvantaged youngsters and concluded: "We have some indication that these programs work, but virtually no information about what works best with different populations" (p. 103). Programs that work well for some groups of children may not be as effective with others. Children from different cultural groups sometimes differ in learning styles. Tutorial techniques may work best for some youngsters, while group experiences may be more effective for others.

CONCLUSIONS

Active parent involvement with program goals may be the crucial component that increases the effectiveness of an intervention curriculum. Sadly, this means that when children live in particularly chaotic situations with drug-addicted families, sometimes with no parents, program efficacy may be lower despite the fine quality of curriculum materials and caregivers. Thus, administrators should not dismiss excellent curricular ideas simply because in a particular case they did not result in long-term sustained intellectual post-program gains. When family supports and sustained

learning experiences in the family can also be implemented, these curricula have proved effective in supporting educational success, particularly for females (Honig, 1982; Lally, Mangione, & Honig, 1988).

Although alternative, more pessimistic, points of view exist, early intervention has been found to increase intellectual achievements. Careful analyses of program effects across models, theoretical bases, target populations, time of entry, length of program participation, and other parameters suggest that program enhancement of preschool intellectual achievements will be more probable. Among these parameters are the following scenarios: Families are initially at high risk for failure to support early learning; enrichment begins in infancy; the ratio of caregivers to children is high in preschool classrooms; practices are developmentally appropriate; curricula emphasize hands-on activities; rich language experiences are carried out by highly trained teachers in stable employment; and parents are actively recruited and supported through in-home visitation as teachers of their own children.

HEAD START

Project Head Start was conceived initially as a proactive governmental attempt to "eradicate the negative effects of poverty on children's development" (Zigler & Styfco, 1993, p. 1). Under the auspices of the United States Department of Health and Human Services, Head Start provides 3-to-5-year-old children from low-income families with half-day enrichment programs to boost socialization and cognitive skills and to detect health problems and facilitate treatment so early learning is not impeded (Zigler & Muenchow, 1992). Thus, Head Start services comprehensively include cognitive, socialization, dental, mental health, medical, language, nutrition, and social services. In 1990, Head Start programs enrolled 575,802 at-risk preschoolers, out of a total of 2,475,000 poor children eligible for the program. During the twenty-fifth anniversary year of the founding of Head Start, nearly one-half million preschoolers in 24,000 Head Start classrooms were daily receiving several hours of cognitive and social boosts to their development (Horn, 1990).

Early, short-range studies showed positive gains in IQ scores and better reading readiness or language test

results. Results from the Westinghouse Learning Corporation and Ohio University study of long-range effects of Head Start revealed, however, that when matched with first, second, and third graders who had not attended the program, the Head Start graduates from 104 centers were still considerably below national norms on standardized tests of language and scholastic achievement. Summer programs alone were ineffective in producing cognitive gains, and full-year programs were only marginally effective. Children who had attended Head Start, however, are more advanced than control children in kindergarten, and they continued to outperform controls in first grade (Lee et al., 1990). The use of IQ as an outcome measure of Head Start and of most early intervention programs does more than measure formal cognitive abstract abilities such as reasoning and speed of information processing. Two researchers have reported that IQ "is also an achievement test highly influenced by the child's particular experiences . . . [and] intelligence test performance is greatly influenced by a variety of motivational and/or personality variables that have little to do with either formal cognition or achievement variables" (Zigler & Trickett, 1978, pp. 792–793). They suggest that IQ changes resulting from early intervention programs reflect primarily motivational changes that influence test performance rather than actual changes in cognitive functioning and propose that social competence (rather than IQ to which it is related in complex ways) be used as "the major measure of the success of intervention programs such as Head Start" (p. 793).

Goals of the Follow Through program proposed by President Lyndon B. Johnson in 1967 were to build on Head Start by providing continuing help to the children as they entered regular school. Because of funding constraints, Follow Through became a limited planned variation experiment with twenty-two sponsors, to determine whether various curricular models would have differential effects on low-income children. In 1969, a limited number of communities participating in Follow Through were invited to join the Head Start Planned Variation Program (HSPV) to enhance continuity. Participants achieved substantially greater test gains than children not enrolled in any preschool but no better results than comparison children in regular Head Start. Goodson, Swartz, and Mill-sap (1991) found that the effectiveness of models

varied from site to site and that no model succeeded everywhere it had been tried. Direct instructional models had a higher average effect by third grade on basic skills compared with some models, but particular HSPV models were not significantly better than others (Rivlin & Timpane, 1975).

In response, Head Start funded thirty demonstration projects in 1986–1987 to develop strategies to aid preschoolers in making the transition from Head Start to kindergarten. These linkages presumably will help disadvantaged youngsters to sustain the gains they make in Head Start. In addition, teacher qualifications have been gradually improved so that 79 percent of Head Start teachers have a Child Development Associate degree or similar credential, and parents are encouraged toward upward educational mobility, which will enhance chances for sustaining children's cognitive achievements (Horn, 1990).

Despite ongoing attempts to improve Head Start and to provide linkages and transition supports for participants as they graduate to public school, no definitive answer exists to the question of whether early intervention programs such as Head Start can actually raise intelligence per se. This uncertainty stems from several circumstances: Head Start children enter intervention after the infant-toddler period, when intellectual growth could be more easily influenced; intervention classrooms differ markedly in developmental appropriateness in implementation; the extent of teacher education in child development and early childhood education is not uniformly high; and the degree of support for substantial, sensitive, and sustained parental involvement in the child's early learning varies widely among families and classrooms. If intellectual success, however, is measured not by IQ gains but by lowered school failure, dropout, and retention rates, then early intervention programs have demonstrated effectiveness in increasing cognitive competence.

(See also: INFANCY; INFANT TESTS AS MEASURES OF EARLY COMPETENCE.)

BIBLIOGRAPHY

- ACHENBACH, T. M., PHARES, V., & HOWELL, C. T. (1990). Seven-year outcome of the Vermont intervention program for low-birthweight infants. *Child Development, 61*, 1672–1681.

- Administration for Children, Youth, and Families (1992, January). *New directions and child and family research: Shaping Head Start in the 90's*. Washington, DC: Administration for Children, Youth, and Families.
- BEREITER, C., & ENGELMANN, S. (1966). *Teaching disadvantaged children in the preschool*. Englewood Cliffs, NJ: Prentice-Hall.
- BERRUETA-CLEMENT, J. R., SCHWEINHART, L. J., BARNETT, W. S., EPSTEIN, A. S., & WEIKART, D. P. (1984). *Changed lives: The effects of the Perry Preschool Program on youths through age 19*. Ypsilanti, MI: High/Scope Press.
- BRYANT, D. M., & RAMEY, C. Y. (1987). An analysis of the effectiveness of early intervention programs for high-risk children. In M. Guralnick & C. Bennett (Eds.), *The effectiveness of early intervention for at-risk and handicapped children*. New York: Academic Press.
- CALDWELL, B. M., WRIGHT, C., HONIG, A. S., & TANNENBAUM, J. (1970). Infant day care and attachment. *American Journal of Orthopsychiatry*, 40, 397-412.
- CLARKE-STEWART, K. A., & FEIN, G. G. (1983). Early childhood programs. In M. M. Haith & J. J. Campos (Eds.), *Handbook of child psychology*, Vol. II: *Infancy and developmental psychology*. New York: Wiley.
- Consortium for Longitudinal Studies. (1983). *As the twig is bent: Lasting effects of preschool programs*. Hillsdale, NJ: Erlbaum.
- DAY, M. C., & PARKER, R. K. (EDS.). (1977). *The preschool in action* (2nd ed.). Boston: Allyn & Bacon.
- FOWLER, W., & KHAN, N. (1974). *The later effects of infant group care: A follow up study*. Toronto, Ontario, Canada: Ontario Institute for Studies in Education.
- GARBER, H. L. (1988). *The Milwaukee Project: Preventing mental retardation in children at risk*. Washington, DC: American Association on Mental Retardation.
- GOODSON, B. D., SWARTZ, J. P., & MILLSAP, M. A. (1991). *Working with families: Promising programs to help parents support young children's learning: Summary of findings*. Cambridge, MA: Abt Associates.
- HONIG, A. S. (1979). *Parent involvement in early childhood education*. Washington, DC: National Association for the Education of Young Children.
- HONIG, A. S. (1982). Intervention strategies to optimize infant development. In E. Aronowitz (Ed.), *Prevention strategies for mental health* (pp. 25-55). New York: Neale Watson.
- HORN, W. F. (1990). Head Start: Facing new challenges. *Children Today*, 19(3), 4-5.
- KEISTER, M. E. (1970). *The 'good life' for infants and toddlers*. Washington, DC: National Association for the Education of Young Children.
- LALLY, J. R., MANGIONE, R., & HONIG, A. S. (1988). The Syracuse University Family Development Research Program: Long-range impact of an early intervention with low income children and their families. In D. Powell (Ed.), *Parent education as early childhood intervention: Emerging directions in theory, research, and practice* (pp. 79-104). Norwood, NJ: Ablex.
- LAZAR, I., DARLINGTON, R., MURRAY, H., ROYCE, J., & SNIPPER, A. (1982). Lasting effects of early education: A report from the consortium for longitudinal studies. *Monographs of the Society for Research in Child Development*, 47(2-3, Serial No. 195).
- LEE, V. E., BROOKS-GUNN, J., SCHNUR, E., & LIAW, F. (1990). Are Head Start effects sustained? A longitudinal follow-up comparison of disadvantaged children attending Head Start, no preschool, and other preschool programs. *Child Development*, 61, 495-507.
- LEVENSTEIN, P. (1988). *Messages from home: The Mother-Child Home Program and the prevention of school disadvantage*. Columbus: State University of Ohio Press.
- MCCARTNEY, K. (1984). Effects of the quality of the daycare environment on children's language development. *Developmental Psychology*, 20, 244-260.
- MCQUEEN, A. B., & WASHINGTON, V. (1990). Effects of intervention on the language facility of poor, black, adolescent mothers and their preschool children. In A. S. Honig (Ed.), *Optimizing early child care and education* (pp. 137-152). London: Gordon & Breach.
- National Council of Jewish Women (1982). *HIPPY International Workshop*. Publication No. 91. Israel: School of Education, Hebrew University of Jerusalem.
- OLDS, D. L., HENDERSON, C. R., CHAMBERLIN, R., & TATELBAUM, R. (1985). Improving the delivery of prenatal care and outcomes of pregnancy: A randomized trial of nurse home visitation. *Pediatrics*, 77, 16-28.
- RAMEY, C., & GOWAN, J. (1990). A general systems approach for modifying risk for retarded development. In A. S. Honig (Ed.), *Risk factors in infancy* (pp. 9-26). London: Gordon & Breach.
- RIVLIN, A. M., & TIMPANE, P. M. (1975). *Planned variation in education: Should we give up or try harder?* Washington, DC: The Brookings Institution.
- RUOPP, R., TRAVERS, J., GLANTZ, F., & COELEN, C. (1979). *Children at the center*. Cambridge, MA: Abt Associates.

- SCHWEINHART, L. J., BARNES, H. V., & WEIKART, D. P. (1993). *Significant benefits: The High/Scope Perry Preschool Study through age 27*. Monograph Number 10. Ypsilanti, MI: High/Scope Press.
- WASIK, B. H., RAMEY, C. T., BRYANT, D. M., & SPARLING, J. J. (1990). A longitudinal study of two early intervention strategies: Project CARE. *Child Development*, 61, 1682–1696.
- ZIGLER, E., & MUENCHOW, S. (1992). *Head Start: The inside story of America's most successful educational experiment*. New York: Basic Books.
- ZIGLER, E., & STYFCO, S. (EDS.). (1993). *Head Start and beyond: A national plan for extended childhood intervention*. New Haven, CT: Yale University Press.
- ZIGLER, E., & TRICKETT, P. K. (1978). IQ, social competence, and evaluation of early childhood intervention programs. *American Psychologists*, 33(9), 789–798.

ALICE STERLING HONIG

INTERVENTIONS, LATER Intellectual competence and performance are determined by several factors. In addition to genetic endowment these factors include acquired knowledge, skills, and attitudes. The acquisition of these assets is a lifelong process. What individuals learn during their formative years is widely acknowledged to influence greatly their cognitive development and performance for a long time. Later experiences also can be important determinants of the intellectual capabilities and functioning of adults.

Although interest among educators and researchers in teaching thinking in school is not new, it has been particularly high in recent years. This interest has been expressed in research on the teaching and learning of thinking, in the development and promotion of ways to teach thinking, in the development of programs and materials intended to facilitate that teaching, in the convening of many conferences addressed to the subject, and in the publication of numerous books and articles that focus on various aspects of thinking and how it can be enhanced.

Reasons for this interest in the teaching of thinking include practical concerns about the anticipated needs of an intellectually-able workforce in the future, the belief that a thinking citizenry is the best assurance of

the survival of democracy, and the philosophical perspective that the ability to think has a lot to do with what it means to be human. Furthermore, there are data that indicate that the educational system has not been doing the best possible job in this respect (National Assessment of Educational Progress, 1981; National Commission on Excellence in Education, 1983). These data provide strong motivation to find more effective ways to teach thinking.

ASPECTS OF EFFECTIVE THINKING

Thinking has many aspects, and there are many ways to attempt to improve it. Researchers and developers of programs to teach thinking have approached the problem from a variety of perspectives, and they have focused their attention on several aspects of what it means to think well. The following review of the work that has been done (from Nickerson, 1988/1989) highlights the aspects of thinking emphasized by researchers and program developers:

1. basic operations or processes (e.g., classification, generalization, deduction)
2. subject-matter knowledge (e.g., knowledge about physics, biology, literature, bicycles, postage stamps)
3. knowledge of formal principles or standards of reasoning (e.g., logic, statistics)
4. knowledge of informal principles and tools of thought (e.g., rhetoric, problem-solving heuristics)
5. metacognitive knowledge (e.g., knowledge about human cognition, knowledge of one's own cognitive strengths and weaknesses, knowledge about how to monitor and control one's own thought processes)
6. values/attitudes/dispositions/styles (e.g., reflectiveness, fairness, objectivity)
7. beliefs (e.g., about problems, approaches, self, causes, importance of effort)

Basic Operations or Processes. Some researchers and program developers have attempted to identify mental operations, or skills, that seem to be involved in high-level thinking across a variety of subject areas (physics, automotive repair, computer programming) and to develop methods for improving them (Ehrenberg & Ehrenberg, 1982; Feuerstein et al., 1980; Marzano et al., 1988). The specific skills identi-

fied by these efforts vary considerably from program to program. What are called skills in some contexts are called abilities, operations, processes, or something else in others. The common idea is that there are a few abilities that are fundamental to effective thinking in any context; the hope is that a focus on these abilities and an attempt to improve them through instruction explicitly designed for that purpose will improve thinking generally.

Subject-matter Knowledge. Some investigators have focused on generally useful skills. Others have emphasized the importance of the acquisition of knowledge of specific subjects and of problem-solving techniques that are especially useful for work in those areas. Those who support this latter view point out that thinking does not occur in a vacuum; to think, one must think *about* something, and the quality of the thinking is likely to depend greatly on how much one knows about the subject about which one is thinking. They note too that expert problem solvers in a particular subject area tend to be distinguished from novices by the fact that they not only know more about the area but that their knowledge is organized in a more useful way (Larkin et al., 1980). Related to the assumption that subject-matter knowledge is essential to good thinking is the assumption that the problem-solving techniques that are most effective in specific contexts tend to be relatively unique to those contexts and that, therefore, the teaching of approaches to problem solving should also be done on a subject-matter basis (Chi, Glaser, & Rees, 1982).

The series of books on U.S. history, entitled *Evaluating Viewpoints: Critical Thinking in United States History Series*, written by a high-school teacher (O'Reilly, 1990–1992), approaches the teaching of thinking within the context of traditional subject matter. The first book of the series begins with a focus on critical thinking as a topic of study in its own right; subsequent units deal with specific aspects of critical thinking—identifying and evaluating sources, evaluating evidence, evaluating arguments—in the context of discussions of historical events. Another example of embedding instruction of thinking within a context other than a traditional elementary or high-school subject is Waller's (1988) text, which uses the setting of the courtroom and a jury trial.

Formal Principles or Standards of Reasoning. The effectiveness of instruction in formal logic to improve the thinking of people who receive such instruction has been debated for a very long time. Much of the debate centers on the question of how relevant logic, at least as traditionally taught, is to the kinds of reasoning required by the challenges of everyday life (Fisher, 1988; Toulmin, 1958). In recent years, considerable attention has been given to the question of how effectively people reason about situations that are characterized by many possible outcomes and a high degree of uncertainty (e.g., a weather forecaster trying to determine the likelihood of rain, a military officer assessing the chances that a particular strategy will be successful, a physician judging the probability that a patient has a specified disease) and to whether reasoning in such situations is improved by training in statistics and mathematical probability theory. There is considerable evidence that people who lack training in statistics and probability often reason inappropriately when they have to judge the likelihoods of uncertain events (Kahneman & Tversky, 1982; Nisbett & Ross, 1980; Tversky & Kahneman, 1974). Although training in statistics and probability theory is no guarantee against such errors, there is some evidence that people who have had such training are less likely to make them than those who do not (Nisbett et al., 1987).

Informal Tools of Thought. Human problem solving has been studied by psychologists almost since the beginning of psychology as an experimental science. Most of the early work was addressed to the question of how people approach problems when left to their own devices; little came out of this work in the nature of *prescriptions* for problem solving. Since the late 1950s, however, there has been considerable interest among psychologists and others in the possibility of specifying generally effective problem-solving procedures. This interest was sparked, in part, by the appearance on the scene of electronic computers. If such procedures could be specified precisely, they could be implemented as computer programs and the usefulness of computers would thereby be enhanced.

A variety of strategies, or "heuristic" principles and procedures, seem to distinguish problem solving that is more effective from problem solving that is less effective. There have been several attempts to isolate

what is known about effective problem-solving strategies, and to organize this information for people who wish to study problem-solving techniques on their own. Heuristic problem-solving techniques have also been included in thinking-skills programs designed for classroom use (Adams, 1986; Sternberg, 1986) as well as the focus of a number of college-level courses (Rubenstein, 1975; Schoenfeld, 1985; Wheeler & Dember, 1979).

Metacognition—Thinking about Thinking.

Several researchers have stressed the importance of metacognitive knowledge, or METACOGNITION more generally, as a factor in thinking. *Metacognition* means knowledge about cognition in general and about one's own cognitive strengths and weaknesses in particular and about the ability to manage and monitor one's own performance on cognitively-demanding tasks. Among the earliest advocates of the teaching of metacognition were Brown (1978, 1980) and Flavell (1979, 1981). Although the teachability of specific strategies—sometimes called *executive control strategies*—to direct and manage one's own thought processes is a matter of continuing debate, some investigators have reported success in this regard (Belmont, Butterfield, & Ferretti, 1982).

Attitudinal Factors. Some efforts to improve thinking have emphasized the importance of attitudes or dispositions and similar factors, such as values and styles (Paul, 1984, 1992; Ennis, 1985, 1987; Lipman, 1991; Newmann, 1991). One can hardly doubt either the beneficial effects of such characteristics as inquisitiveness, fairmindedness, and reflectiveness, or the negative impact of their opposites. The question of how to teach or promote attitudes, dispositions, intellectual values or styles that are conducive to good thinking remains a challenge for research. Investigators who have stressed such factors have typically stressed also the importance of supportive environments, that is, communities of learning or inquiry that encourage students to think and share the products of their thinking and that reward them for doing so.

Beliefs. Closely associated with the emphasis on attitudes and dispositions is the idea that beliefs also can influence the quality of one's thinking. One can hold any of a variety of beliefs about the nature of knowledge, about the roles of effort and luck in

achievement, about the modifiability of one's own capabilities and limitations, about the usefulness of learning what is taught in school, about personal responsibility in opinion formation and decision making, and so on. What one believes about these and similar issues can affect not only how one thinks but how much effort one puts into thinking (Dweck & Eliot, 1983; Baron, 1985; Deci & Ryan, 1985).

These emphases are different, but they are not mutually incompatible; although investigators differ considerably with respect to where they believe the emphasis should be placed, most would acknowledge the importance of aspects of thinking in addition to those they stress in their own work. Most would probably agree that there are learnable skills and strategies that can be applied in many contexts. Few would question the importance of subject-specific knowledge. No one would deny that the quality of thinking is influenced by attitudes, dispositions, and beliefs. The question of exactly how best to improve thinking through instruction and education is much debated, but the idea that the quality of thinking is influenced by many factors is not.

ACADEMIC LEVEL OF PROGRAMS

Approaches to the teaching of thinking have been developed for applications at all levels of education. Some of the better-known programs, such as Instrumental Enrichment (Feuerstein et al., 1980) and Philosophy for Children (Lipman, Sharp, & Oscanyan, 1980) are intended for use primarily (though not necessarily exclusively) in lower- and middle-school settings. Others, are more appropriate for upper-middle or high-school contexts.

The fact that many high-school graduates arrive at college without the thinking ability that the college curriculum assumes (Mullis & Jenkins, 1988) has motivated the development of a variety of programs intended to strengthen the thinking skills of incoming college students who need such remediation. Examples include ADAPT (Accent on the Development of Abstract Processes of Thought) (Fuller et al., 1980), COMPAS (Consortium for Operating and Managing Programs for the Advancement of Skills) (Schermerhorn, Williams, & Dickison, 1982), and SOAR (Stress On Analytical Reasoning) (Carmichael et al., 1980).

Much of the guidance that has come out of the work on problem-solving strategies and metacognition is applicable to people of any age, but in particular to those whose school years are behind them, as well as to those who are still in the process of acquiring their formal education. This guidance is available not so much in formal programs as in books that are accessible to any interested reader.

EVIDENCE OF EFFECTIVENESS

A large amount of material has been published regarding the teaching of thinking. Many approaches have been developed either to teach thinking in school contexts or to help people improve their thinking on their own. There are two types of evidence that these approaches work and result in their intended improvements in thinking. One is the results-oriented documentation that would be produced by formal evaluations, that is, objective data showing what a program can accomplish. Another is a rationale that rests on a well-developed and tested theory of thinking or specific aspects of thinking.

Some approaches have been subjected to evaluative testing, and a few have some theoretical basis. Some have neither claim to credibility. The latter represent their developers' beliefs about how thinking should be taught, but they provide no evidence of effectiveness beyond that. This is not to say that all such approaches are wrong, but only to caution that in the absence of either empirical evidence of effectiveness or a compelling theoretical rationale, claims of what a program will accomplish should be considered as only predictions that need to be put to the test.

Fair and conclusive evaluations of educational innovations are very difficult to conduct. In any educational system there are variables that are beyond an investigator's control. As a consequence, the data from even the most carefully conducted evaluation experiments often are open to more than one interpretation, and, therefore, do not provide unequivocal answers to the questions to which the evaluations were addressed. Reliable evaluative data are relatively scarce, partly because of how difficult it is to conduct educational evaluations and perhaps because program developers have, in some cases, been sufficiently convinced of the merits

of their own approaches that they have not been highly motivated to attempt evaluations.

Although most program developers are careful not to make unwarranted claims for what their programs can be expected to accomplish, some of the descriptions of commercially available material are more promotional than is appropriate from a strictly scientific point of view. A few program developers or authors do make unsubstantiated and highly dubious claims of being able to effect substantial increases in general intellectual competence through the use of certain mental exercises over a relatively short period of time. Although the line between approaches that have sound scientific bases and those that do not is not as easy to draw as one might wish, the following questions are worth raising when attempting to judge the probable merits of a program or approach that one is considering adopting.

1. What are the program's specific objectives? Are they clear? In what ways is the program intended to enhance thinking? Are its objectives the ones you wish to pursue?
2. Is the program's documentation clear and adequate? Does it provide the guidance needed to use the program effectively?
3. What evidence exists that the program's objectives are realized when the program is used as intended? What evidence is there that any improvements that the program effects generalize and last? In the absence of such evidence, what arguments are made that the anticipated improvements should do so, and how compelling are they?
4. Are the assumptions regarding teacher qualifications clear? Be cautious if no assumptions are made in this regard or if the teacher's qualifications are treated as unimportant. The teacher's own thinking competence is a critical factor in the utilization of any structured effort to enhance thinking through instruction.
5. Does the promotional literature make the program sound too good to be true? If so, it probably is. Thinking is complex, and its many aspects are only partially understood. Improving it is possible, but it is an ambitious undertaking and there are no effective, quick and easy ways of doing so.

FURTHER READING

Reviews of research on the teaching of thinking include Resnick (1987) and Nickerson, Perkins, and Smith (1985). Overviews of many research and demonstration programs can be found in Chipman, Segal, and Glaser (1985), Costa (1985), Baron and Sternberg (1986), Schwebel and Maher (1986), Presseisen (1988), and Mulcahy, Short, and Andrews (1991). In addition to edited collections of descriptive overviews of programs, there are books that give extensive information on specific programs. Representative of such books are *Instrumental Enrichment* (Feuerstein et al., 1980), *Philosophy in the Classroom* (Lipman, Sharp, & Oscanyan, 1980), and *Intelligence Applied* (Sternberg, 1986). Swartz and Perkins (1989) have provided an overview of work in the field written explicitly for teachers and other educational practitioners.

Several published sets of program materials are intended for classroom use. These include *Basics* (Ehrenberg & Ehrenberg, 1982), *Odyssey* (Adams, 1986) and *Tactics for Thinking* (Marzano & Arredondo, 1986). Information relevant to the teaching, or learning, of thinking skills is to be found also in numerous books that deal with one or another aspect of what it means to think effectively and suggest ways to improve one's own thinking (Adams, 1974; Bransford & Stein, 1984; Chaffee, 1990; Halpern, 1989; Hayes, 1989; Jones & Idol, 1990; Kahane, 1984; Nickerson, 1986; Rubenstein, 1975; Ruggiero, 1984; Wickelgren, 1974). Some of these books emphasize creative thinking, some emphasize critical thinking, some emphasize problem solving or what might be called pragmatic thinking.

BIBLIOGRAPHY

- ADAMS, J. L. (1974). *Conceptual blockbusting: A guide to better ideas*. San Francisco: W. H. Freeman.
- ADAMS, M. J. (COORDINATOR). (1986). *Odyssey: A curriculum for thinking*. Watertown, MA: Mastery Education Corporation.
- BARON, J. (1985). *Rationality and intelligence*. New York: Cambridge University Press.
- BARON, J., & STERNBERG, R. J. (EDS.). (1986). *Teaching thinking skills: Theory and practice*. New York: W. H. Freeman.
- BELMONT, J. M., BUTTERFIELD, E. C., & FERRETTI, R. P. (1982). To secure transfer of training instruct self-management skills. In D. K. Detterman & R. J. Sternberg (Eds.), *How and how much can intelligence be increased?* Norwood, NJ: Ablex.
- BRANSFORD, J. D., & STEIN, B. S. (1984). *The IDEAL problem solver: A guide for improving thinking, learning, and creativity*. New York: W. H. Freeman.
- BROWN, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology*. Hillsdale, NJ: Erlbaum.
- BROWN, A. L. (1980). Metacognitive development and reading. In R. J. Spiro, B. C. Bruce, & W. F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, NJ: Erlbaum.
- CARMICHAEL, J. W., HASSELL, J., HUNTER, J., JONES, L., RYAN, M., & VINCENT, H. (1980). Project SOAR. *The American Biology Teacher*, 42(3), 169–173.
- CHAFFEE, J. (1990). *Thinking critically* (3rd ed.). Boston: Houghton-Mifflin.
- CHI, M., GLASER, R., & REES, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 1). Hillsdale, NJ: Erlbaum.
- CHIPMAN, S. F., SEGAL, J. W., & GLASER, R. (EDS.). (1985). *Thinking and learning skills: Vol. 2. Research and open questions*. Hillsdale, NJ: Erlbaum.
- COSTA, A. L. (ED.). (1985). *Developing minds: A resource for teaching thinking*. Alexandria, VA: Association for Supervision and Curriculum Development.
- CUBAN, L. (1984). Policy and research dilemmas in the teaching of reasoning: Unplanned designs. *Review of Educational Research*, 54, 655–681.
- DECI, E. L., & RYAN, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum.
- DWECK, C. S., & ELIOT, E. S. (1983). Achievement motivation. In P. H. Mussen (Ed.), *Handbook of child psychology 4*. New York: Wiley.
- EHRENBERG, S. D., & EHRENBERG, L. M. (1982). *BASICS: Building and applying strategies for intellectual competencies in students*. Coshocton, OH: Institute for Curriculum and Instruction.
- ENNIS, R. H. (1985). Critical thinking and the curriculum. *National Forum*, 65, 28–31.
- ENNIS, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron & R. J. Sternberg (Eds.), *Teaching thinking skills*. New York: W. H. Freeman.
- FEUERSTEIN, R., RAND, Y., HOFFMAN, M., & MILLER, R.

- (1980). *Instrumental enrichment*. Baltimore, MD: University Park Press.
- FISHER, A. (1988). *The logic of real arguments*. New York: Cambridge University Press.
- FLAVELL, R. H. (1979). Metacognition and cognitive monitoring: A new area of psychological inquiry. *American Psychologist*, 34, 906–911.
- FLAVELL, R. H. (1981). Cognitive monitoring. In W. P. Dickson (Ed.), *Children's oral communication skills*. New York: Academic Press.
- FULLER, R. G., BERGSTROM, R. F., CARPENTER, E. T., CORZINE, H. J., MCSHANE, J. A., MILLER, D. W., MOSHMAN, D. S., NARVESON, R. D., PETR, J. L., THORNTON, M. S., & WILLIAMS, V. G. (EDS.). (1980). *Piagetian approaches in higher education*. Lincoln, NE: ADAPT Program.
- HALPERN, D. F. (1989). *Thought and knowledge: An introduction to critical thinking* (2nd ed.). Hillsdale, NJ: Erlbaum.
- HAYES, J. R. (1989). *The complete problem solver* (2nd ed.). Hillsdale, NJ: Erlbaum.
- JONES, B. F., & IDOL, L. (1990). *Dimensions of thinking and cognitive instruction*. Hillsdale, NJ: Erlbaum.
- KAHANE, H. (1984). *Logic and contemporary rhetoric: The use of reason in everyday life* (4th ed.). Belmont, CA: Wadsworth.
- KAHNEMAN, D., & TVERSKY, A. (1982). On the study of statistical intuitions. *Cognition*, 11, 123–141.
- LARKIN, J. H., MCDERMOTT, J., SIMON, D. P., & SIMON, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208, 1335–1342.
- LIPMAN, M. (1991). *Thinking in education*. New York: Cambridge University Press.
- LIPMAN, M., SHARP, A. M., & OSCANYAN, F. *Philosophy in the classroom*. Philadelphia: Temple University Press.
- MARZANO, R. J., & ARREDONDO, D. E. (1986). *Tactics for thinking*. Aurora, CO: Mideast Regional Educational Laboratory.
- MARZANO, R. J., BRANDT, R. S., HUGHES, C. S., JONES, B. F., PRESSEISEN, B. Z., RANKIN, S. C., & SUHOR, C. (1988). *Dimensions of thinking: A framework for curriculum and instruction*. Alexandria, VA: Association for Supervision and Curriculum Development.
- MULCAHY, R. F., SHORT, R. H., & ANDREWS, J. (EDS.). (1991). *Enhancing learning and thinking*. New York: Praeger.
- MULLIS, I. V. S., & JENKINS, L. B. (1988). *The science report card: Elements of risk and recovery*. Princeton, NJ: Educational Testing Service.
- NATIONAL ASSESSMENT OF EDUCATIONAL PROGRESS. (1981). *Reading, thinking, and writing: Results from the 1979–1980 national assessment of reading and literature* (Report No. 11-L-01). Denver, CO: Education Commission of the States.
- NATIONAL COMMISSION ON EXCELLENCE IN EDUCATION. (1983). *A nation at risk: The imperative for educational reform*. Washington, DC: U.S. Government Printing Office.
- NEWMANN, F. M. (1991). Higher order thinking in the teaching of social studies: Connections between theory and practice. In J. F. Voss, D. N. Perkins, & J. W. Segal (Eds.), *Informal reasoning and education* (pp. 381–400). Hillsdale, NJ: Erlbaum.
- NICKERSON, R. S. (1986). *Reflections on reasoning*. Hillsdale, NJ: Erlbaum.
- NICKERSON, R. S. (1988/1989). On improving thinking through instruction. In E. Z. Rothkopf (Ed.), *Review of research in education*, Vol. 15 (pp. 3–58). Washington, DC: American Educational Research Association.
- NICKERSON, R. S., PERKINS, D. N., & SMITH, E. E. (1985). *The teaching of thinking*. Hillsdale, NJ: Erlbaum.
- NISBETT, R. E., & ROSS, L. (1980). *Human inference: Strategies and shortcomings of social judgment*. Englewood Cliffs, NJ: Prentice-Hall.
- NISBETT, R. F., FONG, G. T., LEHMAN, D. R., & CHENG, P. W. (1987). Teaching reasoning. *Science*, 238, 625–631.
- O'REILLY, K. (1990–1992). *Evaluating viewpoints: Critical thinking in United States history series* (4 vols.). Pacific Grove, CA: Midwest Publications.
- PAUL, R. (1992). *Critical thinking: What every person needs to know in a rapidly changing world*. Sonoma, CA: Foundation for Critical Thinking.
- PAUL, R. W. (1984). Critical thinking: Fundamental to education for a free society. *Educational leadership*, 42, 4–14.
- PRESSEISEN, B. Z. (1986). *Critical thinking and thinking skills: State of the art definitions and practice in public schools*. Philadelphia: Research for Better Schools.
- PRESSEISEN, B. Z. (ED.). (1988). *At-risk students and thinking: Perspectives from research*. Washington, DC: National Education Association.
- RESNICK, L. B. (1987). *Education and learning to think*. Washington, DC: National Academy Press.
- RUBENSTEIN, M. F. (1975). *Patterns of problem solving*. Englewood Cliffs, NJ: Prentice-Hall.
- RUGGIERO, V. R. (1984). *The art of thinking: A guide to critical and creative thought*. New York: Harper & Row.
- SCHERMERHORN, L. L., WILLIAMS, L. D., & DICKISON, A. K. (1982). *Project COMPAS: A design for change*. Sanford, FL: Seminole Community College.

- SCHOENFELD, A. (1985). *Mathematical problem solving*. New York: Academic Press.
- SCHWEBEL, M., & MAHER, C. A. (EDS.). (1986). *Facilitating cognitive development: International perspectives, programs, and practices*. New York: Haworth Press.
- STERNBERG, R. (1986). *Intelligence applied: Understanding and increasing your intellectual skills*. New York: Harcourt Brace Jovanovich.
- SWARTZ, R. J., & PERKINS, D. N. (1989). *Teaching thinking: Issues and approaches*. Pacific Grove, CA: Midwest Publications.
- TOULMIN, S. (1958). *The uses of argument*. New York: Cambridge University Press.
- TVERSKY, A., & KAHNEMAN, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, 185, 1124–1131.
- WALLER, B. N. (1988). *Critical thinking: Consider the verdict*. Englewood Cliffs, NJ: Prentice Hall.
- WHEELER, D. D., & DEMBER, W. N. (EDS.). (1979). *A practicum in thinking*. Cincinnati, OH: University of Cincinnati.
- WICKELGREN, W. A. (1974). *How to solve problems*. San Francisco: W. H. Freeman.

RAYMOND S. NICKERSON

INTUITION Intuition was originally a philosophical concept implying the immediate grasping of an object or idea without any apparent conscious reasoning. Intuition is thus a close cousin of sudden insight. In psychology, however, intuition has fared considerably less well than insight. The different reception accorded these closely related notions probably derives from subtle differences in the everyday usage of these two terms.

In common parlance, *intuition* implies a feeling that a particular solution, decision, guess, or premonition is correct, even if it flies in the face of evidence to the contrary. An intuition in this sense may occasionally prove justified, but such an outcome is typically regarded as rare and therefore attributable to blind luck or uncommon genius. *Insight* implies that a problem has been successfully resolved in light of available evidence. So, even if the solution turns out to be mistaken, it is viewed as an understandable error, rationally made. Further, intuition but not insight is often viewed as antithetical to analytic thinking, which is regarded as the conscious or at least potentially con-

scious use of abstract symbols, critically and rationally employed in accordance with rules of logic. Indeed, intuition is often defined, at least in part, as nonanalytic thought. By way of contrast, there is no danger of self-contradiction in asserting that analytic thinking about a problem can yield a sudden insight into its solution. Finally, because of the first two points, intuition is often tainted by an implication of irrationality, illogicality, or mental laziness, whereas this is less likely to be the case for insight.

INTUITIVE JUDGMENTS ARE ERROR-PRONE

A great deal of research on decision making and judgment has further diminished the tenuous reputation of intuition. These investigations, spearheaded by Amos Tversky and Daniel Kahneman (1983), have shown that people are very prone to make erroneous decisions on the basis of intuition. One example helps to demonstrate their point.

People were asked for their intuitive judgments about how many seven-letter words ending in “ing” would occur in four pages of a novel. A second question asked how many words would have “n” as the sixth of seven letters. Notice that all words ending in “ing” comprise a subset of words having “n” as the sixth letter, so it stands to reason that “----n-” words will be far more frequent than “---ing” words. In fact, however, people estimated that there were almost three times more words ending in “ing” than words having “n” as the sixth letter.

How could people’s intuitions have been so far off the mark? As it happens, people find it much easier to think of specific words ending in “ing” than non-“ing” words having “n” as the sixth letter. It therefore seems probable that the relatively high accessibility of seven-letter words ending in “ing” unduly influenced estimates of their frequency. In fact, a great deal of research has shown that intuitive judgments are biased by factors such as how readily information comes to mind—unexamined for its plausibility or pertinence in light of other background considerations. This research also indicates that humans are almost perverse in their tendency to make such intuitive decisions. It has been argued that this tendency compromises human rationality because it so often generates avoidable

errors in judgment. What is more, being well educated does not necessarily prevent a person from falling into such “intuition traps.” Even very sophisticated people are prone to make such errors of intuition, which—once pointed out—may seem obvious and foolish to them.

INTUITION AND PROBLEM SOLVING

Other work related to intuition is less pessimistic in its implications. Consider an early investigation of problem solving reported by Norman Maier. People in this study were confronted with the following problem: Tie together the free ends of two strings that are suspended from the ceiling, but so far apart from each other that a person cannot grasp both strings at the same time, even if the person grasps one and walks it toward the other. One elegant solution to this problem involved realizing that a pair of pliers, which were laying in full view, could be used as a pendulum bob. Thus, the pliers could be tied to the end of one string and swung toward the second one. The person could then go to the second string, grasp it, and await the arrival of the first one. It was then simply a matter of liberating the first string from the pliers and tying it to the second string.

Thirty-seven people who did not solve the problem in about ten minutes were given an important clue, though it was not announced as such. The clue involved the experimenter “inadvertently” nudging one of the strings, thereby putting it into lateral motion. After being exposed to this clue, twenty-three people solved the problem—most of them within sixty seconds. The vast majority of these people had no idea that the movement of the string had inspired the solution, however; in fact, most of them said they hadn’t even noticed the string move. Incidentally, when the experimenter simply twirled the string between his thumb and finger, it did not generate the pendulum-bob solution. So, it seems likely that the *lateral* movement of the string was the key factor in producing the correct solution, even though most people exposed to this clue seemed oblivious to it.

The string-pendulum investigation usefully introduces and illustrates a view of intuition that can be distinguished from both the original philosophical notion of intuition as the immediate grasping of

something true and the late-twentieth-century psychological view of intuition as generating irrational, error-prone judgments.

INTUITION: A REFORMULATION

The above investigation has two quite different implications. On one hand, the findings can be interpreted as revealing a crucial flaw in human cognition: People are unable to provide an informed account of their own behavior, and such ignorance can foster irrational conduct and decisions. This implication squares with the outcome of the word-estimating study mentioned earlier. Suppose people had realized that their intuitive judgments were in danger of being unduly influenced by the relative ease with which they could think of words ending in “ing.” Surely they would have taken this knowledge into account and saved themselves a potentially embarrassing error.

On the other hand, findings from the pendulum-string problem have a positive implication that is not so well revealed in the word-estimating task. It is well established that people can be conscious of only a few things at once. The participants in Maier’s study were tacitly informed by the lateral motion of the string, however, thereby permitting them to transcend the limitations that conscious experience imposed on their efforts to solve the problem. In sum, the possibility that human judgment can be unconsciously *biased* is balanced by the fact that it can also be tacitly *informed* by considerations that are not well represented in conscious experience.

This latter possibility is the basis for an emerging view of intuition that is consistent with late-twentieth-century research in cognitive psychology. According to this view, intuition is not a mental pipeline to the truth. Rather, it involves the automatic activation of relevant information in a manner that can yield helpful insights or hunches.

A basic assumption of this reformulation is that clues to a problem or puzzle reflect and eventually reveal an underlying structure or coherence. If relevant memories and ideas are clue-activated enough to become conscious, their significance for the problem at hand is recognized and experienced as an insight or hunch. Accordingly, we can define intuition as the perception of clues to coherence that tacitly activates and

guides thought toward an insight or hunch about the nature of the coherence in question. Notice that, by this definition, intuition is a mental process that can generate insight as a product.

Figure 1 helps illustrate this reformulated notion of intuition. It is an example of a so-called Gestalt-closure task—in this case, a schematic rendering of a whistle. People who see the whistle often take a few moments to do so. What happens between the time when they see a confusing array of black patches and lines and when they see them as a whistle? According to the above formulation, there is enough “whistleness” in the Gestalt to activate incipient thoughts of a whistle (instead of a trombone, say). If and when the idea of a whistle is activated sufficiently to become conscious, its relevance to the Gestalt is apt to be recognized immediately in the form of a sudden insight.

Implications and Limitations of a Reconceptualized Concept of Intuition. The above formulation has some important implications and limitations. First, intuition is constrained by a person’s past experience. Clues to coherence cannot automatically activate relevant memories unless they are available to be activated. For example, people who have never seen a whistle are unlikely to recognize one when it is depicted in a highly schematic, Gestalt-closure fashion. For such people, there is nothing in their storehouse of memories that can be activated by Gestalt’s “whistleness”; consequently, the “whistleness” of the Gestalt can have no clue value for them. In sum, stored memories and knowledge limit the range over which intuition can operate.

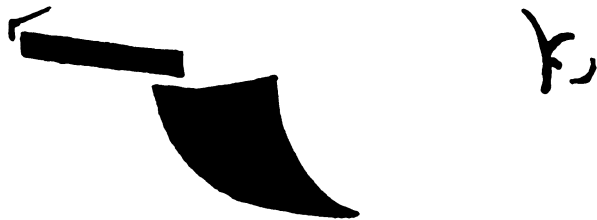


Figure 1
A Gestalt-closure depiction of a whistle

Second, thought is intrinsically intuitive in nature: Thinking inevitably proceeds from partial information to interpretation—from clues to coherence to a hunch about the coherence in question. To borrow Jerome Bruner’s (1955) felicitous phrase, we all go “beyond the information given” (p. 41). However, some people need fewer clues to discern the coherence that underlies them, or discern coherence faster on the basis of the same clues that are available to everyone else. This success is partly a matter of familiarity with a particular domain of inquiry and partly dependent on a variety of other considerations. For instance, in the context of problem solving, a high level of anxiety tends to yield conventional or stereotyped responses, whereas a low to moderate level of arousal has a higher probability of generating productive hunches. In any case, people who seem to have a singular facility for going beyond the information given may be widely regarded by their peers as intuitive. According to the present, reformulated notion of intuition, however, this is an honorific rather than a technical use of the term.

Third, the question of what actually transpires between perceiving clues and perceiving the coherence they reveal remains to be worked out in detail. Some authorities argue that the cognitive processes involved are essentially very fast versions of conscious, analytic processes that have become routinized and habitual through repeated practice. Others suggest that intuition involves fundamentally different cognitive processes from those involved in conscious reasoning. According to this view, intuitive processes are not conducted in words or abstract symbols and are not constrained by logic. Rather, they are triggered more or less automatically by clues in accordance with psychological laws of association and conditioning. Still others emphasize that analytic and intuitive modes of thought form a continuum and that the nature of a problem—and the manner in which it is presented—determine whether an analytic or intuitive approach to problem solving is apt to be more successful. This view emphasizes that an intuitive approach to problem solving is not always inferior to an analytic approach, frequent claims to the contrary notwithstanding.

Finally, during the course of problem solving, even though a hunch or insight may surface suddenly as an “aha experience,” the intuitive processes underlying its

emergence are more likely to be continuous than discontinuous in nature. In other words, sudden insight is not as spontaneous as it seems. Nevertheless, it can be very difficult for people to provide an accurate account of how they got from the clues of a problem to a satisfactory conclusion. This difficulty occurs in part because the sudden emergence of a hunch or solution immediately arrests attention and decreases the likelihood that the person will retrospectively scan the immediately preceding mental processes. In addition, the mental processes that generate hunches are typically implicit and automatic, which makes it intrinsically difficult, if not impossible, for them to be well represented in conscious experience. Thus, systematic investigation of intuition and insight in problem solving will ultimately have to depend on more than the retrospections of the successful problem solver. Attempts in this direction began to appear in the literature in the mid-1980s. One program of research led by Nobel laureate Herbert SIMON (1986) has attempted, with some success, to generate computer simulations of intuition and productive thought.

INTUITION, NEW KNOWLEDGE, AND ERROR

Intuition (along with the somewhat related notion of creativity) addresses the problem of how people generate and acquire new knowledge. This is a philosophical issue that goes back at least to Plato. His theory of anamnesis argued that new knowledge was generated by remembering ideas that the person's soul had known in a previous life. While this notion may seem quaint and outdated, its emphasis on memory is quite contemporary. We do not literally remember solutions to problems, however. Rather, the clues provided by a problem activate relevant memories, and then we *recognize* how they can be adapted to the problem at hand. Moreover, this process does not guarantee a correct answer. To the contrary, the possibility for generating new knowledge guarantees the inevitability of error. Indeed, intuitive processes that could not generate error could not generate anything new at all, including genuine insights into the nature of things.

(See also: INSIGHT.)

BIBLIOGRAPHY

- BASTICK, T. (1982). *Intuition: How we think and act*. New York: Wiley.
- BOWERS, K. S., REGEHR, G., BALTHAZARD, C., & PARKER, K. (1990). Intuition in the context of discovery. *Cognitive Psychology*, 22, 72–110.
- BRADSHAW, G. F., LANGELY, P. W., & SIMON, H. A. (1983). Studying scientific discovery by computer simulation. *Science*, 222, 971–975.
- BRAINERD, C. J., & REYNA, V. F. (1990). Gist is grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, 10, 3–47.
- BRUNER, J. (1955). Going beyond the information given. In H. E. Gruber, K. H. Hammond, & R. Jessor (Eds.), *Cognition: The Colorado symposium* (pp. 41–69). Cambridge, MA: Harvard University Press.
- DREYFUS, H. L., & DREYFUS, S. E. (1986). *Mind over machine: The power of human intuition and expertise in the era of the computer*. New York: Free Press.
- HAMMOND, K. R., HAMM, R. M., GRASSIA, J., & PEARSON, T. (1987). Direct comparison of the efficacy of intuitive and analytical cognition in expert judgment. *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-17, 753–770.
- KAHNEMAN, D., SLOVIC, P., & TVERSKY, A. (EDS.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.
- KOESTLER, A. (1964). *The act of creation: A study of the conscious and unconscious in science and art*. New York: Dell.
- METCALFE, J., & WIEBE, D. (1987). Intuition in insight and noninsight problem solving. *Memory and Cognition*, 15, 238–246.
- PERKINS, D. N. (1981). *The mind's best work*. Cambridge, MA: Harvard University Press. A very readable introduction to productive thinking.
- PLATT, W., & BAKER, R. A. (1969). The relation of the scientific "hunch" to research. *Journal of Chemistry Education*, 8, 1969–2002.
- POLANYI, M. (1969). The logic of tacit inference. In M. Grene (Ed.), *Knowing and being: Essays by Michael Polanyi* (pp. 138–158), Chicago: University of Chicago Press. (Originally published in *Philosophy*, 1966, 41, 1–18.)
- REBER, A. (1989). Implicit learning and tacit knowledge. *Journal of Experimental Psychology: General*, 118, 219–235.
- ROOT-BERNSTEIN, R. S. (1988). Setting the stage for discovery. *The Sciences*, May/June, 26–35. A good introduction to scientific discovery.

- SHEPERD, R. N. (1978). Externalization of mental images and the act of creation. In B. J. Randhawa & W. E. Coffman (Eds.), *Visual learning, thinking and communication* (pp. 133–189). New York: Academic Press.
- SIMON, H. A. (1986). The information processing explanation of gestalt phenomena. *Computers in Human Behavior*, 2, 241–255.
- SIMONTON, D. K. (1980). Intuition and analysis: A predictive and explanatory model. *Genetic Psychology Monographs*, 102, 3–40.
- STERNBERG, R. J. (1988). *The nature of creativity: Contemporary psychological perspectives*. New York: Cambridge University Press.
- TVERSKY, A., & KAHNEMAN, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review*, 90, 293–315.
- WEISBERG, R. W., & ALBA, J. W. (1981). Gestalt theory, insight, and past experience: Reply to Dominowski. *Journal of Experimental Psychology: General*, 110, 193–198.
- WESTCOTT, M. R. (1968). *Toward a contemporary psychology of intuition: A historical, theoretical, and empirical inquiry*. New York: Holt, Rinehart, & Winston.

KENNETH S. BOWERS

IQ GAINS OVER TIME In twenty countries, every one for which data exist, each generation out-scores the previous generation on IQ tests, often by huge margins. These IQ gains over time probably began no later than the last decade of the nineteenth century at a time when, paradoxically, IQ tests did not yet exist. The twenty countries are: Britain, Northern Ireland, Canada, the United States, New Zealand, and Australia; Norway, Sweden, and Denmark; France, Belgium, and the Netherlands; the former East and West Germanies, Austria, and Switzerland; Israel; Brazil; and China and Japan.

MAGNITUDE OF IQ GAINS

IQ gains covering the last sixty years of the twentieth century reveal a pattern. The gains are largest on tests with culture-reduced content such as RAVEN PROGRESSIVE MATRICES, many of which are indicators of fluid intelligence (*G_f*) (see FLUID AND CRYSTALLIZED INTELLIGENCE, THEORY OF). These tests assume a mini-

um of knowledge and aim at measuring on-the-spot problem-solving ability. They vary widely in content: Tests of spatial visualization, figural similarities and sequences, figure classification and generalization are the purest measures of fluid intelligence; tests such as number series and verbal analogies are less clear measures. The very best data, primarily military tests of comprehensive samples of young men, show Belgium, the Netherlands, and Israel gaining at a rate of 20 points over a generation (thirty years), while Norway, Sweden, and Denmark have gained at a rate of about 10 points (Emanuelsson & Svensson, 1990; Flynn, 1987; Teasdale & Owen, 1989).

The best data for culture-reduced tests suggest an average gain of about 15 points (or one standard deviation) per generation. The high quality of these data means we must take the national differences they reveal seriously. Clearly, Scandinavian countries have gained at a much lower rate than other nations; indeed, beginning in 1985, Sweden may have become the first nation to begin to register losses on a test of spatial visualization. The next best data tend to put Britain, Australia, and Canada, nations similar in culture, at 12 to 16 points per generation (Flynn, 1987, pp. 176, 180; Raven, Court & Raven, 1992, pp. G22–G26). The British estimate is based on adults; samples of schoolchildren give much lower rates. Weak data show wide-ranging results for France, the Germanies, Northern Ireland, China, Brazil, and New Zealand.

Other kinds of tests all measure crystallized intelligence (*G_c*) to some degree. These tests place less emphasis on on-the-spot problem solving and more on whether someone has acquired the skills, or general knowledge, or vocabulary we would expect an intelligent person to gain in a normal life. IQ gains over time diminish as tests get farther and farther from measuring fluid intelligence. The continuum from fluid to crystallized runs from tests with culture-reduced content, through performance tests, through verbal tests, to pure vocabulary tests. Verbal tests always involve spoken or written language. Performance tests involve an operation, using blocks to copy a design, arranging pictures in a logical sequence.

Wechsler performance scale gains among schoolchildren show much the same range of results as the culture-reduced tests they resemble: Only the weaker data give gains much above 20 points over a generation

(thirty years) or below 9 points (Flynn, 1987, pp. 185–186). However, tests like Raven's have generated much adult data, the Wechsler tests very little. There is no obvious tendency for gains to diminish with age but recent Japanese data, from a small sample, suggest caution. Japanese schoolchildren have doubled the rate of gain of white American children, about 20 points compared to 9, while Japanese and U.S. adults show similar rates.

Verbal IQ gains vary from almost nil to 20 points per generation, with 9 as a rough median, and some of this is adult data from military testing. Among the eleven countries that allow a comparison, there is not one in which verbal gains match the gains on culture-reduced, or performance, or nonverbal tests and often the ratios run against verbal gains by two or three to one. Where vocabulary gains can be distinguished from verbal gains in general, they rarely match them. Despite sizable IQ gains in Scotland and Northern Ireland, there have been no vocabulary gains. British adults of all ages gained 27 points over fifty years on Raven's, but gained only 6 points over forty-five years on the Mill Hill Vocabulary Scale (Flynn, 1987, pp. 185–186; Flynn, 1990, p. 47; Lynn, 1990, p. 139; Raven et al., 1992, pp. G22–G26).

PRACTICAL IMPLICATIONS

IQ gains over time mean that IQ tests must be restandardized frequently, otherwise subjects are being scored against obsolete norms and will get inflated IQ scores. For example, when the Wechsler Intelligence Scale for Children (WISC) was restandardized in 1989, the new test manual showed that average persons might be scoring as high as 107 to 108 on the performance scale of the old test, which would make them appear well above average. We can of course try to use past IQ gains to estimate what has happened since a test was last standardized, but if trends alter, we go astray. Using past trends it would have been reasonable in 1989 to assume that a score of 105 was average on both the old WISC performance and verbal scales. The new manual suggests that 107.4 was needed on the former and only 102.4 on the latter (see also WECHSLER SCALES OF INTELLIGENCE).

Obsolete tests play havoc with both group comparisons and diagnosis of the potential of individuals. Ver-

non, back in 1982, did not realize that many studies had scored Chinese Americans against obsolete norms. They were not being compared to whites of their own generation, whose IQs were much the same, but rather to whites of an earlier generation, which made them appear to be an elite IQ group. This obscured an important group difference concerning IQ and achievement: Chinese Americans outperform white Americans academically and in terms of occupational status by huge margins *despite* having no higher mean IQ (Flynn, 1991). Another example, Raven Progressive Matrices, was normed on adults in urban China in 1986. Because there were no up-to-date norms for British adults, it seemed possible that Chinese were equal or even superior. The British norms of 1992 show that the Chinese mean is almost surely below the British mean (Raven et al., 1992, pp. G22–G26).

Intervention and adoption studies go astray because group comparisons are implicit. When Heber and Garber used a massive enrichment program (the Milwaukee Project) to help ghetto blacks, they appeared to have lifted the IQs of children, perhaps destined for mental retardation, 20 points above the average score of contemporary whites. Thanks to obsolete tests, they were comparing these black children to the far lower scores of whites a generation ago. The Skodak and Skeels Adoption Study found a 20-point gap between the IQ of adopted children and their biological mothers, which seemed to show the potency of being raised in a good adoptive home. However, some of that gap was due to norms fourteen years out of date, which means that the adoptive homes were getting credit for what were merely IQ gains over time enjoyed by all children (Flynn, 1984, pp. 40–41).

A widely held hypothesis has been that as people age they suffer large losses in fluid intelligence. This results from comparing today's 70-year-olds with today's 20-year-olds, which gives the elderly much lower scores. When today's 70-year-olds are compared with the 20-year-olds of fifty years before, their actual cohorts, losses with age are seen to be minimal. The apparent "losses" are primarily one generation being outscored by the next generation (Raven et al., 1992, pp. G22–G26). This may seem to show that the elderly of today are at least less intelligent than the youth of today. Even this is suspect because, as we shall see, IQ differences between generations may not constitute in-

telligence differences. Sometimes it is children whose potential is mismeasured. A schoolchild may be scored against an obsolete IQ test and an up-to-date achievement test. Therefore, although actually average for both, he or she appears to be achieving below his or her intellectual potential because of an inflated IQ score.

Finally, a sobering phenomenon. Between 1947 and 1972, a gap of twenty-five years between standardizations of the WISC, IQ gains over time lowered the number of American children eligible to be classified as mentally retarded from 8.8 million to only 2.6 million. Either the criterion was far too strict in 1947 or far too lenient in 1972. There is no evidence in the literature that clinical psychologists were aware of the discrepancy (Flynn, 1985).

IQ GAINS AND INTELLIGENCE

IQ gains over time pose a problem in that some gains seem too large to equate with intelligence gains. This assertion presumes an operational concept of intelligence, however rough and ready. The concept presumed will be labeled "understanding-baseball intelligence." It is derived from an account Arthur JENSEN gives to illustrate the limitations of a subject with a Wechsler IQ of 75. Despite the fact that the man in question volunteered baseball as his chief interest, and attended or viewed games frequently, he was vague about the rules, did not know how many players comprised a team, could not name the teams his home team played, and could not name any of the most famous players (Jensen, 1981, p. 65).

In 1992, John Raven restandardized the Raven Progressive Matrices on a representative sample of the adult population of Dumfries in Scotland, selected as typical of an area whose norms matched those of Britain as a whole (J. Raven, 1981, p. RS1.25). His father, J. C. Raven, had originally normed the test in 1942. For ages 20 to 30, he selected soldiers in army camps whose education matched that of British males at the time. For older ages, he tested large samples from a private firm and from a government department, a majority of whose employees joined as youths and remained until retirement. They gave him a curve of performance from one age cohort to another and he grafted that curve onto his military sample, thereby

deriving norms covering all ages (Foulds & Raven, 1948; J. C. Raven, 1941).

The 1992 sample shows that by that date performance on Raven's peaked no earlier than ages 35 to 40. The 1942 sample peaked earlier but maintained top performance until 35 to 40, so those ages allow the fairest comparison: they also match the results of comparing all ages from 18 to 67. The 1992 sample outscored the 1942 sample by almost 27 IQ points. Moreover, the cohort aged 70 in 1992 matched the 20-year-olds from 1942. If we assume that the 70-year-olds of 1942 matched the 20-year-olds of 1892, we can trace IQ gains back 100 years, that is, back before IQ tests existed. The Raven data are like rays of light from a distant star that give us a glimpse of the universe as it was long before the earth was born (Raven et al., 1992, pp. G22-G26).

The data show that Britons gained 55 IQ points between 1892 and 1992. This means that in terms of today's norms, adult Britons had an average IQ of 45 in 1892. The samples do not match the quality of military samples, and the equating of age cohorts from present to past can only be approximate. Nonetheless it will be difficult to defend any estimate that puts mean IQ in 1892 above 60; therefore, at a minimum, 84 percent had an IQ below 75. In order to identify IQ gains with understanding-baseball intelligence, we would have to assume that 84 percent of Britons could not, even if it became their chief interest, understand cricket in 1892. The best data available pose similar problems. Using military data, Dutch males in 1952 had a mean IQ of 80 in terms of 1982 norms. Can we assume that almost 40 percent of them lacked the capacity to understand soccer, their most favored national sport?

These scenarios are derived from gains on tests of fluid intelligence. Jensen's subject had a Wechsler IQ of 75, and Wechsler tests measure a mix of fluid and crystallized intelligence. It may be said that such tests are a better measure of mental retardation and, therefore, that our scenarios are suspect. This demands a reply on two levels.

First, there are the U.S. data from tests that measure a mix of fluid and crystallized intelligence. Wechsler and Stanford-Binet samples show gains at an average rate of about .3 points per year from 1932 to 1989. However, these gains probably began no later

than 1918. Every study from that era shows large gains, and they are supported by comparing performance on the Stanford-Binet by soldiers in 1918 with the standardization sample of 1932. The evidence suggests at least 21 points gained between 1918 and 1989 (Flynn, 1984). This means that in 1918, when scored against today's norms, Americans had an average IQ of 79 on tests whose crystallized component is at least as great as Wechsler tests. Does that mean that 40 percent of Americans lacked the capacity to understand the basic rules of baseball?

Second, theory posits a functional relationship between fluid and crystallized intelligence such that their problems cannot be compartmentalized: What afflicts one must affect the other. A population whose fluid intelligence tests at 60 or 70 or 80 should not soar much above that for crystallized intelligence. It is quite possible that subjects whose fluid intelligence did not decline until old age should retain the information and vocabulary they acquired earlier, at a time when their fluid intelligence was normal. The evidence of many studies suggests this (Horn, 1989). However, it is quite another thing to imagine people acquiring normal levels of knowledge and vocabulary whose fluid intelligence never, during their entire lives, rose much above the level of mental retardation. Indeed, the very fact that massive IQ gains can occur, whether measured by tests of fluid or crystallized intelligence, despite nil or minimal vocabulary gains seems incongruous.

Theory also gives tests of fluid intelligence an indispensable role in drawing the distinction between intelligence and learning. As an approach to a pure test of learning, imagine a simple task such as tying one's shoes: Since almost everyone can perform this task, it comes close to measuring no intelligence differences. Tests like the Wechsler tests measure a mix of learning and intelligence, that is, the tasks are of graduated difficulty so that they rank people in terms of the amount of intelligence usually required to learn them. Tests of fluid intelligence are designed to presuppose a minimum of learning, often no more than acquaintance with certain simple shapes, and therefore are deemed approaches to a pure test of intelligence (Horn, 1989; Jensen, 1979, pp. 79–81; Jensen, 1980, pp. 234–236).

Jensen and others are now experimenting with replacing IQ tests, at least when they compare groups with results that seem suspect as intelligence differ-

ences, with physiological measures. The hope is that by measuring the electrical response of the cerebral cortex to sights and sounds, how quickly people can move from one button to another, and the time taken for an injection of glucose to reach and be absorbed by the brain, we will find an intelligence test that always gives plausible results, even when used to compare generations over time (Jensen, 1988, 1989).

THE TWO CAUSAL PROBLEMS

Massive IQ gains pose two causal problems and, therefore, candidates for the role of cause must meet two basic criteria: They must be capable of explaining the gains when confronted with the total array of data, and they must be capable of causing IQ gains without at the same time causing intelligence gains.

There are also a few supplementary criteria: They must be environmental and go beyond mere increased exposure to IQ tests. Over one or two generations, only a fanatic eugenics program could have made a significant contribution to IQ gains, and, if anything, mating trends have been dysgenic. The gains antedate the period when testing was common and have persisted into the era when IQ testing, due to its unpopularity, has become less frequent. Even when naive subjects are repeatedly exposed to a variety of tests, IQ scores rise by only 5 or 6 points and the rate of gain reduces sharply after the first few exposures. IQ trends over time show gains as high as 55 points and in some countries, the rate of gain has risen decade after decade.

Education recommends itself as a possible cause. In most countries larger numbers of people are spending longer periods of their life being schooled and examined on academic subject matter. IQ gains in Denmark appear highly correlated with increased years of schooling and more people attaining higher credentials (Teasdale & Owen, 1987). However, the reverse is true in the Netherlands, where matching across generations to hold educational level constant eliminates only 6.5 percent of a massive gain (Flynn, 1987, p. 188). Gains among schoolchildren, say, comparing sixth or twelfth graders with their counterparts a generation ago, cannot be influenced by years of schooling in that the number of years are by definition the same. As for quality of schooling, the tendency of IQ gains to es-

calate the farther one gets from tests correlated with traditional academic skills is virtually universal. The United States has had significant IQ gains since 1972, a period during which the National Assessment of Educational Progress (the nation's "report card") found little or no academic achievement gains.

The fact that education cannot explain IQ gains as an international phenomenon does not, of course, disqualify it as a dominant cause at a certain place and time. Particular countries are sometimes influenced by a factor that is culture specific. Comparing age cohorts suggests that urban China gained 22 IQ points on the Raven Progressive Matrices between 1936 and 1986 (Raven & Court, 1989, p. RS4.8). Learning to read Chinese characters involves endless practice in breaking a complex pattern into two components, one conveying meaning, the other pronunciation. The spread of literacy might be a dominant cause of matrices gains peculiar to urban China.

It is logically possible that peculiar factors dominate in each and every one of our twenty countries. But this seems highly unlikely. Two attempts to explain IQ gains as an international phenomenon have gained currency, namely, the Brand hypothesis and the nutrition hypothesis.

Brand argues that the permissive society advantages the present generation on tests with time limits. The scrupulous test taker of the past wasted time trying to get every item correct; today's subjects are prone to intelligent guessing and finish more items within the time allotted. This hypothesis is theoretically ideal. It explains IQ gains in terms of something that implies no intelligence gains and cites environmental factors independent of mere exposure to tests. However, it has now been falsified. Wechsler performance tests in nation after nation show gains as high as the culture-reduced tests even though the latter usually impose much greater time pressure (Flynn, 1987, pp. 185–186). The huge gains of adult Britons on Raven's were made by subjects who took the test untimed. Flieller, Jautz, and Kop (1989, pp. 11–12) analyzed a Binet-type test with a fairly even balance of timed and untimed items. They found that the last generation left more questions unanswered on both kinds of items, and that worse performance on items completed accounted for virtually all of the last generation's score deficit.

Unlike Brand's hypothesis, the nutrition hypothesis creates a tension between our two basic causal problems. Better nourished brains would function better in the test room, but they should also function better in everyday life. Therefore, if nutrition has caused 20 or 30 or 50 points of IQ gains, we seem driven to posit huge understanding-baseball intelligence gains.

Richard Lynn (1987, 1989) enhances its plausibility by ascribing only 15 points to nutrition and the remainder to other causes such as defective tests. For example, the Raven Progressive Matrices are held to measure increased arithmetical skills as well as intelligence gains and therefore to overestimate intelligence gains. The critique of Raven's poses many evidential problems. Norwegian draftees made Raven gains while suffering losses on a test modeled on the Wechsler adult arithmetic subtest. Military samples from Israel show comparable male and female performance on the matrices, which runs counter to most gender data on mathematics. As for the magnitude of Raven gains, they are larger than those of other nonverbal tests in Britain, but equivalent in Belgium and smaller in Australia, Canada, and Scotland (Flynn, 1987, pp. 173–174, 176; Flynn, 1990).

Even if given a diminished explanatory role, the nutrition hypothesis has its own peculiar evidential problems. Lynn (1987, p. 467) cites a one standard deviation (SD) height gain over the last fifty years, which equals his estimate of British intelligence gains over that period. However, some European countries have been making height gains for fully a century or two and these amount to more than one SD, sometimes to two or three (Floud, Wachter, & Gregory, 1990, pp. 16, 23, 26). If height gains are truly accompanied by intelligence gains, they pose a familiar question: Did the Dutch in 1864 really have the same intelligence as people who today score 65 on IQ tests? Did Norwegians in 1761 really resemble those who today score 62?

The best experimental study of the effects of vitamin-mineral supplements on IQ shows that in California a modest supplement had little effect, a moderate one had significant effect, and a large one had little effect (Schoenthaler et al., 1991, pp. 357–358). That every nation has continuously enhanced nutrition just the right amount, neither too little nor too much, for decade after decade, seems unlikely. For example, the

Netherlands almost certainly gave children born after World War II better nutrition than it gave those born during the great war-time famine. The effect on IQ gains over time was nonexistent (Flynn, 1992, p. 346).

The experimental data from dietary supplements also show that 75 percent of subjects enjoy very modest gains, while 25 percent, presumably subjects who are subclinically malnourished, make large gains. The latter tend to have lower IQs than the former, which means that if enhanced nutrition is a factor, IQ gains over time should come disproportionately from those with below-average IQs. Denmark fits that pattern, but most nations do not. A good sign that IQ gains extend to every IQ level is that score variance remains unchanged over time, or diminishes only because of clear ceiling effects. Military samples or samples of equivalent quality show this for Belgium, Norway, Sweden, Israel (males), Canada, and New Zealand. Dutch Raven data and U.S. Wechsler data actually provide the full IQ curves and allow us to trace gains at all levels (Bouvier, 1969, pp. 4–5; Clarke, Nyberg, & Worth, 1978, p. 130; Elley, 1969, p. 145; Emanuelsson & Svensson, 1990; Flynn, 1985, p. 240; Rist, 1982, p. 47; Teasdale & Owen, 1989, pp. 258–259).

FUTURE OF AN ILLUSION

Massive IQ gains over time present us with something that is both deceptive and real. People today cannot have much more understanding-baseball intelligence than people did generations ago. However, they really are better at some sort of problem-solving ability that must have subtle effects on everyday life. It is as if juggling ability had dramatically escalated over the last 100 years, but no one had noticed until recently, no one could provide a general explanation, and no one could find evidence of enhanced performance at a culturally significant sport, such as baseball or cricket. Some day sufficient data may allow us to see why some tests show higher gains than others, why some periods show higher gains than others, why some countries show higher gains than others, and we will agree on a solution to the causal problems. The result will be a better theory of intelligence and better vehicles for measuring it.

(See also: INTERVENTIONS, LATER; STAGES OF COGNITIVE DEVELOPMENT.)

BIBLIOGRAPHY

- BOUVIER, U. (1969). *Evolution des cotes à quelques tests* [Evolution of scores from several tests]. Brussels: Belgian Armed Forces, Center for Research Into Human Traits.
- CLARKE, S. C. T., NYBERG, V., & WORTH, W. H. (1978). *Technical report on Edmonton Grade III achievement: 1956–1977 comparisons*. Edmonton, Canada: University of Alberta.
- ELLEY, W. B. (1969). Changes in mental ability in New Zealand schoolchildren. *New Zealand Journal of Educational Studies*, 4, 140–155.
- EMANUELSSON, I., & SVENSSON, A. (1990). Changes in intelligence over a quarter of a century. *Scandinavian Journal of Educational Research*, 34, 171–187.
- FLIELLER, A., JAUTZ, M., & KOP, J.-L. (1989). Les réponses au test mosaïque à quarante ans d'intervalle [Answers on the mosaic test after an interval of forty years]. *Enfance*, 42, 7–22.
- FLOUD, R., WACHTER, K., & GREGORY, A. (1990). *Height, health and history*. Cambridge: Cambridge University Press.
- FLYNN, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, 95, 29–51.
- FLYNN, J. R. (1985). Wechsler intelligence tests: Do we really have a criterion of mental retardation? *American Journal of Mental Deficiency*, 90, 236–244.
- FLYNN, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, 101, 171–191.
- FLYNN, J. R. (1990). Massive IQ gains on the Scottish WISC: Evidence against Brand et al.'s hypothesis. *The Irish Journal of Psychology*, 11, 41–51.
- FLYNN, J. R. (1991). *Asian Americans: Achievement beyond IQ*. Hillsdale, NJ: Erlbaum.
- FLYNN, J. R. (1992). Cultural distance and the limitations of IQ. In J. Lynch, C. Modgil, & S. Modgil (Eds.), *Education for cultural diversity: convergence and divergence* (pp. 343–360). London: Falmer Press.
- FOULDS, G. A., & RAVEN, J. C. (1948). Normal changes in the mental abilities of adults as age advances. *Journal of Mental Science*, 94, 133–142.
- HORN, J. L. (1989). Cognitive diversity: A framework for learning. In P. L. Ackerman, R. J. Sternberg, & R. Glaser (Eds.), *Learning and individual differences: Advances in theory and research* (pp. 61–114). New York: W. H. Freeman.
- JENSEN, A. R. (1979). The nature of intelligence and its relation to learning. *Journal of Research and Development in Education*, 12, 79–95.

- JENSEN, A. R. (1980). *Bias in mental testing*. London: Methuen.
- JENSEN, A. R. (1981). *Straight talk about mental tests*. New York: Free Press.
- JENSEN, A. R. (1988). *The Oddman-Out Test: A chronometric anchor for psychometric tests of g*. Unpublished paper.
- JENSEN, A. R. (1989). Raising IQ without increasing g ? [A review of *The Milwaukee Project: Preventing mental retardation in children at risk*]. *Developmental Review*, 9, 234–258.
- LYNN, R. (1987). Japan: Land of the rising IQ. A reply to Flynn. *Bulletin of the British Psychological Society*, 40, 464–468.
- LYNN, R. (1989). Positive correlation between height, head size and IQ: A nutrition theory of the secular increases in intelligence. *British Journal of Educational Psychology*, 59, 372–377.
- LYNN, R. (1990). Differential rates of secular increase of five major primary abilities. *Social Biology*, 38, 137–141.
- RAVEN, J. (1981). *Manual for Raven's Progressive Matrices and Mill Hill Vocabulary Scales* (research supplement No. 1). London: H.K. Lewis.
- RAVEN, J., & COURT, J. H. (1989). *Manual for Raven's Progressive Matrices and Vocabulary Scales* (research supplement No. 4). London: H.K. Lewis.
- RAVEN, J. C. (1941). Standardization of Progressive Matrices. *British Journal of Medical Psychology*, 19, 137–150.
- RAVEN, J. C., COURT, J. H., & RAVEN, J. (1992). *Manual for Raven's Progressive Matrices and Vocabulary Scales* (section 3). Oxford: Oxford Psychologists Press.
- RIST, T. (1982). *Det intellektuelle prestasjonsnivaet i befolkningen sett i lys av den samfunns-messige utviklinga* [The level of the intellectual performance of the population seen in the light of developments in the community]. Oslo: Norwegian Armed Forces Psychology Service.
- SCHOENTHALER, S. J., AMOS, S. P., EYSENCK, H. J., PERITZ, E., & YUDKIN, J. (1991). Controlled trial of vitamin-mineral supplementation: Effects on intelligence and performance. *Personality and Individual Differences*, 12, 351–362.
- TEASDALE, T. W., & OWEN, D. R. (1987). National secular trends in intelligence: A twenty-year cross-sectional study. *Nature*, 325, 119–121.
- TEASDALE, T. W., & OWEN, D. R. (1989). Continued secular increases in intelligence and a stable prevalence of high intelligence levels. *Intelligence*, 13, 255–262.
- TERMAN, L. M., & MERRILL, M. A. (1937). *Measuring intelligence*. London: George G. Harrap.
- YERKES, R. M. (1921). Psychological examining in the United States Army. *Memoirs of the National Academy of Sciences*, Vol. 15. Washington, DC: U.S. Government Printing Office.

JAMES R. FLYNN

ITEM RESPONSE THEORY See LATENT TRAIT THEORY.