

What's Wrong with Psychology Anyway?

David T. Lykken

When I was an undergraduate at Minnesota in 1949, the most exciting course I took was Clinical Psychology, open to seniors and graduate students and taught by the dynamic young star of the psychology faculty, Paul Everett Meehl. In 1956, back at Minnesota after a postdoctoral year in England, the first course I ever tried to teach was that same one; Meehl, now Chair of the Department, wanted more time for other pursuits. Like most new professors of my acquaintance, I was innocent of either training or experience in college teaching, and I shall never forget the trepidation with which I took over what had been (but, alas, did not long remain) the most popular course in the psychology curriculum.

Years later, Paul asked me to contribute a few lectures to a new graduate course he had created called Philosophical Psychology. Sitting in class that first year, I experienced again the magic of a master teacher at work. Meehl's varied and extraordinary gifts coalesce in the classroom—the penetrating intellect, astonishing erudition, the nearly infallible memory, the wit and intellectual enthusiasm, the conjurer's ability to pluck the perfect illustration from thin air . . . I recall one class that ended late while Paul finished explaining some abstruse philosophical concept called the Ramsey Sentence. I have long since forgotten what a Ramsey Sentence is, and I doubt if fifty people in the world besides Paul and, perhaps, Ramsey himself think the concept is exciting. But Meehl had those students on the edge of their seats, unwilling to leave until they had it whole.

The present paper is a distillation of the three lectures I have been contributing to Paul's Philosophical Psychology. I offer it here in fond respect for the man who has been my teacher and friend for nearly forty years.

I shall argue the following theses:

- (I) Psychology isn't doing very well as a scientific discipline and something seems to be wrong somewhere.

- (II) This is due partly to the fact that psychology is simply harder than physics or chemistry, and for a variety of reasons. One interesting reason is that people differ structurally from one another and, to that extent, cannot be understood in terms of the same theory since theories are guesses about structure.
- (III) But the problems of psychology are also due in part to a defect in our research tradition; our students are carefully taught to behave in the same obfuscating, self-deluding, pettifogging ways that (some of) their teachers have employed.

Having made this diagnosis, I will suggest some home remedies, some ways in which the next generation could pull up its socks and do better than its predecessors have done. Along the way I shall argue that research is overvalued in the Academy and that graduate students should not permit themselves to be bullied into feeling bad about the fact that most of them will never do any worthwhile research.

For reasons that escape me, students have said that they tend to find these illuminating discussions depressing in some way. The first lecture, focusing on the defects of the research tradition, is a particular downer, so I'm told. I think this attitude is shortsighted. By taking a frank look at ourselves and making an honest assessment of our symptoms and defects, it is possible, I think, to see some of the apparent and correctable reasons for these problems.

I. Something Is Wrong with the Research Tradition in Psychology

It is instructive to attempt to follow the progress of a research idea from its germination in the mind of a psychological scientist until it finally flowers (if it ever does) within the pages of an archival journal. If the budding idea seems to its parent to be really promising, the almost invariable first step is to write it up in the form of a grant application directed most commonly to one of the federal agencies. Writing grant applications is laborious and time-consuming, and there is no doubt that many research ideas begin to seem less viable during the course of this process and are aborted at this early stage.

A. Most Grant Applications Are Bad

Applications directed to the National Institute of Mental Health are routed to an appropriate Research Review committee consisting of 10 or 12 established investigators with broadly similar interests who meet for several days three times each year to consider submissions and make recommendations for funding. Although all committee members are nominally expected to read the entire set of applica-

tions (and a few probably do this), the review committees depend largely on the reports of those two or three members who have been assigned principal responsibility for the given proposal. The Institute gets good value from these peer review committees whose members, not wishing to appear foolish or uninformed before their peers at the tri-annual meetings, invest many (uncompensated) hours before each meeting studying their assigned subset of applications and composing well-considered critiques and recommendations. At the meetings, proposals are carefully discussed and evaluated before the committee votes. Of all the applications received by NIMH in a given year, only about 25% are considered promising enough to be actually funded.

B. Most Manuscripts Submitted to the Journals Are Bad

Archival scientific journals also depend upon the peer review system. The editors of most psychological journals do a preliminary screening, returning at once those manuscripts that are the most obviously unacceptable, and then send out the remainder to two or more referees selected for their expertise on the topic of the given paper. Like most academic psychologists of my advanced age, I have refereed hundreds of papers for some 20 journals over the years and can attest that it is a dispiriting business. My reviews tended to be heavily burdened with sarcasm evoked by the resentment I felt in having to spend several hours of my time explicating the defects of a paper which one could see in the first ten minutes' reading had no hope of contributing to the sum of human knowledge. I became troubled by the fact that it was possible for me thus to assault the author's *amour propre* from the safety of the traditional anonymity of journal referees, and I began to sign my reviews and have done so unflinchingly these past 15 years or so. While I continue to be critical, I find that I am very careful to be sure of the grounds for my comments, knowing that the author will know who is talking. It seems to me, in this age of accountability, that authors *ought* to know who has said what about their work and, moreover, that journal readers ought to be able to learn in a footnote which editor or reviewers decided that any given article should have been published.

In any case, whether the reviews are signed or not, the effect of this peer review process is that from 60 to 90% of articles submitted to journals published by the American Psychological Association are rejected.

C. Most Actually Published Research Is Bad

In their 1970 *Annual Review* chapter on Memory and Verbal Learning, Tulving and Madigan reported that they had independently rated each of 540 published articles in terms of its "contribution to knowledge." With "remarkable agreement," they found that they had sorted two-thirds of the articles into a category labeled:

“utterly inconsequential.” The primary function these papers serve is to give something to do to people who count papers instead of reading them. Future research and understanding of verbal learning and memory would not be affected at all if none of the papers in this category had seen the light of day. (Tulving & Madigan, 1970, p. 441)

About 25 percent of the articles were classified as:

“run-of-the-mill” . . . these articles also do not add anything really new to knowledge . . . [such articles] make one wish that at least some writers, faced with the decision of whether to publish or perish, should have seriously considered the latter alternative. (p. 442)

Only about 10 percent of the entire set of published papers received the modest compliment of being classified as “worthwhile.” Given that memory and verbal learning was then a popular and relatively ‘hard’ area of psychological research, attracting some of the brightest students, this is a devastating assessment of the end product.

Hence, of the research ideas generated by these psychologists, who are all card-carrying scientists and who liked these ideas well enough to invest weeks or months of their lives working on them, less than 25% of 40% of 10% = 1% actually appear to make some sort of contribution to the discipline.

D. Most Published Articles Are Not Read Anyway

Garvey and Griffith (1963) found that about half the papers published in APA journals have fewer than 200 readers (not all of whom are edified). Two-thirds of these papers are never cited by another author. Somewhat surprisingly, the same thing is true even in physics: Cole and Cole (1972) found that half the papers in physics journals are never cited. Even articles in *Physical Review*, generally considered one of the most prestigious journals, do not always make much of a splash; 50% are cited once or never during the three years after they appear.

When he was at Minnesota years ago, B. F. Skinner used to say that he avoided reading the literature since it only “poisons the mind.” In psychology, what other researchers are doing is seldom useful to one’s self except perhaps as something to refute or, more rarely, as a bandwagon to climb up on. One does not have to actually read the literature until it is time to start writing one’s paper. Lindsey (1978) and Watson (1982) have cited the long publication lags typical of social science journals as evidence that psychologists do not need to know what their colleagues are doing; we do not fear being ‘scooped’ because it is so unlikely that anyone else would be prospecting in the same area.

E. Theories in Psychology Are Like Old Soldiers: They Are Not Refuted or Replaced—They Don't Die—They Only Fade Away

Like a good scientific theory, this simile of Paul Meehl's has sufficient verisimilitude to continue to be useful. The exciting theoretical developments of my student days—the work of Hull, Spence, and Tolman, to focus just on one then-active area—have sunk into obscurity. In the hard sciences, each generation stands upon the shoulders of its predecessors, the bones of the Elder Giants become part of the foundation of an ever-growing edifice. The great names of psychology's comparatively recent past are respected mainly as intrepid explorers who came back empty-handed. There is no edifice, just this year's ant hill, most of which will be abandoned and washed away in another season.

In the 1940s and '50s, there was a torrent of interest and research surrounding the debate between the S-R reinforcement theorists at Yale and Iowa City and the S-S expectancy theorists headquartered at Berkeley. As is usual in these affairs, the two sides produced not only differing theoretical interpretations but also different empirical findings from their rat laboratories, differences that ultimately led Marshall Jones to wonder if the researchers in Iowa and California might not be working with genetically different animals. Jones obtained samples of rats from the two colonies and tested them in the simple runway situation. Sure enough, when running time was plotted against trial number, the two strains showed little overlap in performance. The Iowa rats put their heads down and streaked for the goal box, while the Berkeley animals dawdled, retraced, investigated, appeared to be making "cognitive maps" just as Tolman always said they did. But by 1965 the torrent of interest in latent-learning had become a backwater and Jones's paper was published obscurely (Jones & Fennel, 1965).

A brilliant series of recent studies of goal-directed behavior in the rat (Rescorla, 1987) demonstrates with elegant controls that the animal not only learns to emit the reinforced response in the presence of the discriminative stimulus but it also learns which response leads to which reward. When one of the reinforcers is devalued (e.g., by associating that type of food pellet with the gastric upset produced by lithium chloride), the rate of that response falls sharply while the animal continues to emit responses associated with different reinforcers. In 1967 these findings would have seemed much more important, embarrassing as they are for the Hull-Spence type of theory. However, in 1987, although these studies were ingenious and produced clear-cut results, they are the results that any layperson might expect and they do not have the surplus value of seeming to contribute to some growing theoretical structure.

The present state of knowledge in psychology is very broad but very shallow. We know a little bit about a lot of things. There are many courses in the psychology curriculum, but few have real prerequisites. One can read most psychology

texts without first taking even an introductory course. But the range or scope of the field is very great so that there will be a majority of people at every APA convention with whom I share few if any scientific interests.

F. Research in Psychology Does Not Tend to Replicate

Charles Darwin once pointed out that, while false theories do relatively little harm, false facts can seriously retard scientific progress. As Mark Twain put it, somewhere, it is not so much what we don't know that hurts us, as those things we do know that aren't so. Weiner and Wechsler (1958), in a similar vein, remark that "the results that are the most difficult to explain are the ones that are not true" (p. ix). Every mature psychologist knows from experience that it is foolish to believe a new result merely on the basis of the first published study, especially if the finding seems unusually important or provocative. Within the narrow circles of our particular fields of interest, many of us learn that there are certain investigators who stand out from the herd because their findings can be trusted.

There is a lot of talk currently about actual dishonesty in research reporting. We were all quite properly scandalized by the Cyril Burt affair when his official biographer concluded that at least most of the subjects in Burt's widely cited study of monozygotic twins reared apart were as fictitious as the two female collaborators whose names Burt signed to his reports of this alleged research (Hearnshaw, 1979). But the problem of the unreplicability of so many findings in the psychological literature involves something more subtle and more difficult to deal with than deliberate chicanery. In almost every study, the investigator will have hoped to find a certain pattern of results, at the very least an orderly, self-consistent pattern of results. The processes of planning, conducting, and analyzing any psychological experiment are complicated, frequently demanding decisions that are so weakly informed by any ancillary theory or established practice as to seem essentially arbitrary. As the investigator makes his or her way through this underbrush, there is the ever-beckoning lure of the desired or expected outcome that tends to influence the choices made at each step.

Selective error-checking is perhaps the simplest and most innocent example of the problem. If the worst sin researchers committed was to re-score or recalculate results that come out 'wrong,' while accepting at once those results that fit with expectations, a significant number of unreplicable findings would appear in the journals. To illustrate some of the subtler sources of distortion, let us consider a couple of real-life examples (see also Gould, 1978).

(1) Marston's Systolic Blood Pressure Lie Detector Test

Before the First World War, psychologist William Moulton Marston discovered what he thought to be Pinocchio's nose, an involuntary physiological reaction that all human beings display when they are deliberately lying but never

when they are telling the truth. Marston's specific lie response was a transitory increase in systolic or peak blood pressure following the (allegedly deceptive) answer. When World War I broke out, the National Research Council appointed a committee to assess the validity of Marston's test as a possible aid in the interrogation of suspected spies. The committee consisted of L. T. Troland of Harvard, H. E. Burtt of Ohio State, and Marston himself. According to Marston (1938), a total of 100 criminal cases were examined in the Boston criminal court and the systolic blood pressure test led to correct determinations in 97 of the 100 cases.

Marston later invented the comic-strip character "Wonder Woman," with her magic lasso that makes men tell the truth. During the 1930s his picture was to be found in full-page magazine advertisements using the lie detector to "prove" that Gillette blades shave closer and more comfortably. For these reasons, we might be skeptical of Marston's scientific claims. But Troland and Burtt were respected psychologists, and Father Walter Summers, chair of the Psychology Department at Fordham, was not a man to be suspected of exaggeration. Summers (1939) invented a lie detector based on an entirely different principle and claimed that his method had proved 100% accurate on a long series of criminal cases. But both Marston and Summers were wrong. Neither method has been taken seriously during the last 50 years and both of the "specific lie responses" they claimed to have discovered are commonly shown by innocent people while truthfully denying false accusations. It is impossible now to discover how it was that the hopes of these enthusiastic investigators became transmuted into false proofs. Their "studies" are not described in detail, the raw data are not available for re-analysis, we do not even know how they established in each case which of the criminal suspects were in fact lying and which were not.

(2) The "Neural Efficiency Analyzer" Scandal

A simple flash of light produces in the brain a complex voltage waveform known as an event related potential (ERP), lasting about half a second after the flash. The ERP can be easily recorded from EEG electrodes attached to the scalp. Because the ERP is weak in comparison with the random background brain-wave activity, a large number of flashes must be presented to obtain an adequate ratio of signal to noise. ERPs to simple stimuli vary in form from person to person but are quite stable over time, and the ERPs of monozygotic twins are very similar in shape. In 1965 John Ertl and William Barry, at the University of Ottawa, reported correlations of $-.88$ and $-.76$ between Wechsler IQ and ERP latency in samples of college students (Barry & Ertl, 1966). If IQ depends primarily upon the speed with which the brain responds to stimulation then, since IQ scores are not perfectly reliable and certainly contain some variance associated with differences in prior learning, a direct, culture-free measure of native intelligence could not be expected to correlate with IQ test scores more strongly than this.

Impressed by this work, the Educational Records Bureau obtained from the Ford Foundation a grant of \$414,000 for a follow-on study (in the 1960s, \$414,000 amounted to real money). The study subjects were 1,000 elementary school children in Mt. Vernon, NY, preschoolers, first- and seventh-graders. At the start of the school year, an ERP was obtained by Ertl's method from each child. In addition, five basic mental abilities were measured by conventional tests. At the end of the year, teacher's ratings, grades, and scores on standardized achievement tests were also collected. The latencies of the various ERP components showed no relationship whatever to any of the intelligence or achievement variables. Hundreds of correlations were computed and they formed a tight Gaussian distribution centered on zero with an SD of about .15 (Davis, 1971).

This large study was a debacle, an utter waste of everybody's time and the Ford Foundation's money, and it should have been avoided. It *would* have been avoided if the team of investigators had included a psychologist trained at Minnesota because he (she) would have been deeply suspicious of those original findings and would have insisted on doing a quick, inexpensive pilot study to find out whether Ertl's remarkable IQ correlations could be replicated in New York.

(3) Perceptual Disorders in Schizophrenia

In 1968, while on sabbatical leave in London, I came upon a remarkable article in the *American Journal of Psychiatry*. A psychiatrist named Bemporad (1967) reported a striking perceptual anomaly in schizophrenic patients. The study was distinguished by Bemporad's exemplary use of separate groups of chronics, acutes (many of these actually tested in the hospital emergency room upon admission), and, most interestingly, a group of previously psychotic patients tested in remission. Thus, one could apparently conclude not only that the phenomenon was not just a consequence of long-term hospitalization but also that it was not merely an effect of psychotic state per se since it appeared almost as strongly among the remitted patients.

Bemporad employed three of the Pseudo-Isochromatic Plates published by the American Optical company and widely used for the assessment of color blindness. These plates are composed of an apparently random pattern of colored dots or circles of various sizes and hues. In each plate a dotted figure or numeral (e.g., "86") can be discerned by a person with normal color vision because the dots making up the figure are of a hue different from the background dots or circles. Because these figural dots or circles are matched for saturation with their neighbors, persons incapable of distinguishing the hues cannot perceive the pattern. Bemporad reasoned that the primitive inability to organize component parts into an integrated perceptual whole, which had been reported for schizophrenics by previous authors, might reflect itself in this test since the perception of the number patterns requires the subject to impose a gestalt upon a set of circles having no common boundary.

Bemporad showed three of the plates, one at a time, to his subjects, asking them only to tell what they saw in each plate. His 20 control subjects made only 2% errors on the three cards, while the chronic, acute, and recovered schizophrenics made 97%, 78%, and 65% errors, respectively.

Because I was currently doing research involving schizophrenic patients at a London hospital, it was easy to arrange a partial replication of the Bemporad study. We thought we might improve slightly on his test simply by using 10 of the pseudo-isochromatic plates, including the 3 that Bemporad employed, and by administering another, easy plate as the first one seen by each subject. The easy plate contained the figure "12" outlined by closely spaced dots that differed both in hue and saturation from the background; it is included in the set as a demonstration plate or as a check for possible malingering. By beginning with this easy sample, we made sure that each subject understood the task (some of these patients, after all, might have been recently shown ink blots and asked, "Tell me what you see.'). We tested 18 schizophrenic patients, some chronic and some in the acute phase of their first admission. We also tested 12 hospital nurses as our control group. All of the subjects were male. The control group was an unnecessary indulgence since we already knew that normal people could see the figures and the only point of our study was to determine whether the Bemporad phenomenon was genuine. British psychiatrists were stricter in their diagnostic practices than American psychiatrists in the 1960s; if the schizophrenic brain had difficulty imposing a gestalt on dotted figures, most of our 18 patients should have made numerous errors on our expanded test.

Our replication required no research grant or fancy preparations. The data were easily collected in a week's time. The results were easily summarized; 29 of the 30 subjects tested correctly identified the figures in all ten plates. The single exception was a patient with specific red-green color blindness who made characteristic errors. While we were never able to account for Bemporad's findings, we could certainly conclude that his empirical generalization was false. This failure to replicate was described in a short note and submitted to the *American Journal of Psychiatry*. After several months, a rejection letter was received from the editor together with an impassioned seven-page critique of my three-page note by an anonymous referee, obviously Bemporad himself.

I then suggested to the editor that it seemed a poor policy to permit an author whose work had failed to replicate to decide whether to publish the report of that failure. The editor agreed and submitted our note to "an altogether neutral referee and a very wise man" who agreed that our study proved the Bemporad phenomenon to be a figment. However, he too recommended against publication of the note on what still seem to me to have been curious grounds. "I doubt whether the readers of the *APA Journal* have even heard of 'Bemporad's phenomenon' any more than I did. . . . So far as I know the original paper has now been forgotten and the new notice which it receives can only give the item new life."

This is, I guess, the “let sleeping dogs lie” principle of editorship and may help account for the fact that, while many psychological research findings do not in fact replicate, comparatively few reports of specific failures to replicate can be found in the journals.

*G. Science Is Supposed to Be a Cumulative Endeavor But
Psychologists Build Mostly Castles in the Sand*

Anyone who reads the recent book *What is Intelligence? Contemporary Viewpoints on its Nature and Definition* edited by Sternberg and Detterman (1986), in which 25 experts responded to the question posed in the title, could easily conclude that there are about as many different conceptions of “intelligence” as the number of experts. This was also true back in 1921 when the same question was asked of an earlier group of experts. Comparing the two symposia, separated by 65 years, we find scarcely more concensus among experts today than in 1921. . . . Shouldn't we expect by now something more satisfying than [this] welter of diverse and contradictory opinions? . . . Where are indications of cumulative gains of research, converging lines of evidence, and generally accepted definitions, concepts, and formulations? (Jensen, 1987, pp. 193–194)

One of the central concepts of psychology—the paradigmatic concept of differential psychology—is intelligence, a topic of great theoretical and practical interest and research for more than a century, the only psychological trait that can boast its own exclusive journal. Yet, in 1987, the leading modern student of intelligence finds it necessary to lament the lack of real cumulative progress in that core area.

Suppose that with some magic Time Machine we could transport Linus Pauling back to the day in 1925 when he had his final oral examination for the Ph.D. in Chemistry at Cal Tech. Our Time Machine will restore his youthful vigor but will permit him to retain all the new things that he has learned, through his own research and that of others, in the 60-plus years since he was examined for his doctorate. Imagine the wonders with which he could regale his astonished professors! Many of the most important developments—the quantum theoretical aspects, for example—would be beyond their understanding. Just a partial description of the technology that is now available in the chemical laboratory would be likely to induce ecstatic seizures in at least some committee members. Those professors of the flapper era would look upon their bright-eyed student as if he were a visitor from some advanced civilization on another planet—as indeed he would be.

Contrast this fantasy now with its psychological equivalent. Let us put Paul Meehl in the Time Machine and send him back to his final oral at Minnesota in 1945. What could he amaze his committee with? What wonders of new technology, what glistening towers of theoretical development, could he parade before their wondering eyes? Shall we tell them the good news about biofeedback? How about the birth and death, without issue, of the Theory of Cognitive Dissonance? What James Olds discovered about pleasure centers in the brain would be exciting, but most of the substantial work that followed would have to be classified as neuroscience rather than psychology.

They will be interested to learn that Hull is dead and that nobody cares anymore about the “latent learning” argument. He could tell them now that most criminals are not helpless victims of neuroses created by rejecting parents; that schizophrenia probably involves a biochemical lesion and is not caused by battle-ax mothers and bad toilet training; that you cannot fully understand something as complex as language by the simple principles that seem to account for the bar-pressing behavior of rats in a Skinner box. In other words, there are some things we know now that many professional psychologists did not know 45 years ago. But it was the professionals who had managed to convince themselves of such odd notions in the first place—their neighbors would have known better. I am sure that each of you could, with some effort, generate a short list of more interesting and solid findings (my own list, not surprisingly, would include some of my own work), but it is a depressing undertaking because one’s list compares so sadly with that of any chemist, physicist, or astronomer. Can we blame it on our youth? Think of the long list of astonishing discoveries produced in our coeval, genetics, with just a fraction of our person-power.

H. Cargo-Cult Science

In his lively autobiography, the late Nobel laureate Richard Feynman (1986) expressed the view that much of psychological research is “Cargo-cult science”:

In the South Seas there is a cargo cult of people. During the war, they saw airplanes land with lots of good materials, and they want the same thing to happen now. So they’ve arranged to make things like runways, to put fires along the sides of the runways, to make a wooden hut for a man to sit on, with two wooden pieces on his head like headphones and bars of bamboo sticking out like antennas—he’s the controller—and they wait for the airplanes to land. They’re doing everything right. The form is perfect. It looks just the way it looked before. But it doesn’t work. No airplanes land. So I call these things cargo cult science, because they follow all the apparent precepts and forms of scientific investigation, but they’re missing something essential, because the planes don’t land.

Summary

It is hard to avoid the conclusion that psychology is a kind of shambling, poor relation of the natural sciences. As the example of genetics shows us, we cannot reasonably use our relative youth as an excuse—and at age 100 we are a little long in the tooth to claim that with a straight face anyway. Psychologists in the American Association for the Advancement of Science have been trying recently to get *Science* to publish a psychological article now and then. The editors reply that they get lots of submissions from psychologists but they just are not as interesting as all the good stuff they keep getting from the biochemists, the space scientists, the astronomers, and the geneticists.

Moreover, *Science*, like its British counterpart, *Nature*, is a relatively fast-publication journal where hot, new findings are published, findings that are of general interest and that other workers in the field will want to know about promptly. But psychologists seldom have anything to show and tell that other psychologists need to know about promptly. We are each working in a different part of the forest, we are not worried that someone else will publish first, and we do not need to know what others have found because ours is not a vertical enterprise, building on what has been discovered previously.

Most of us realize that we do not really have to dig into the journals until we are ready to write up our own work for publication and need some citations to make our results seem more relevant and coherent. Our theories have a short half-life and they just die in the larval stage instead of metamorphosing into something better. Worse yet, our experiments do not replicate very well and so it is hard to be sure what to theorize about.

II. Why? What Has Gone Wrong with Psychology's Research Tradition?

A. Are Psychologists Dumber Than Physicists or Biologists?

Many years ago W. S. Miller administered the advanced form of his Analogies test to graduate students in various disciplines at the University of Minnesota. Ph.D candidates in psychology ranked with those in physics and math and higher than those in most other fields. Graduate Record Examination scores of students applying now for graduate work in psychology are still very high.

Every now and then an eminent 'hard' scientist decides to devote his later years to fixing up psychology. Donald Glaser, who won a Nobel Prize for inventing the bubble-chamber, became a psychologist and sank into obscurity. More recently, Crick, of Double Helix fame, has started theorizing about the function of dreams. I predict that the Freudian theory will outlive the Crickian. We are probably not actually dumber than scientists in the more progressive disciplines

(I wish I could really be sure of this), and it seems doubtful that the problems of psychology can be attributed to a failure to attract bright young researchers. One cannot be sure how long that is going to hold true (or even if it's true now) because the competition for really bright, energetic young minds is fierce.

B. Psychology Is More Difficult, More Intractable, Than Other Disciplines

(1) It Is Hard to See the Forest for the Trees

Everybody is at least an amateur psychologist since we all exemplify the system in question and we each need to understand and predict behavior, our own and that of others; for most of us this imperative is stronger than our need to understand the genes or the stars. This constant intimacy with the raw material of our science is often helpful in the sense of doing armchair experiments or as a source of ideas but, on balance, it is more of a hindrance. Scientists must be able to idealize and oversimplify, to escape from the particular to the general and, often, to sneak up on a complicated problem by successive approximations. The atomic model of Thompson and Bohr, for example, served physics well for many years and probably made possible the new knowledge and the new concepts which ultimately proved that model to be grossly oversimplified. If Thompson and Bohr had known some of what is now known about leptons and quarks and so on, if they had been required to operate in the murk of all these forbidding complexities, they might never have been able to make any progress. The same thing is true in biology. It was important for nineteenth-century biologists to be able to think of the cell as a simple little basic unit of protoplasm with relatively simple functions and properties. If they had been forced to realize that the average cell does more complicated chemistry every day than the Dupont Corporation, it might have been very inhibiting.

When one looks at the heavens on a clear night, it is interesting to contemplate the fact that only a few hundred stars are visible to the naked eye at any given time and place, only about 6,000 in the entire celestial sphere. Moreover, only a few really bright stars are present and they combine in our perception as the great constellations. The constancy of shape of these starry patterns and their regular apparent movement from east to west was the beginning of astronomy. If we had had the eyes of eagles, there would have been millions of visible stars in the night sky, perhaps too many for us to be able to distinguish clear patterns. The north star, Polaris, essential to the ancient navigators, is easily located by any child; the lip of the Big Dipper points it out. Could a child with eagle's eyes find Polaris so easily with hundreds of distractor stars visible in the intervening space which seems empty to the human eye? Now we speak in a familiar way about island universes in their billions, each containing billions of suns, about pulsars and quasars and black holes. It is possible that these great achievements of human

understanding would have been impeded and delayed if our vision had been clearer so that the true complexity of the heavens had been more thrust upon us.

Good scientists need to be capable of a kind of tunnel vision, to be able to ignore even obvious difficulties long enough for their vulnerable newborn ideas to mature sufficiently to be able to survive on their own. This is difficult for psychologists because we live inside an exemplar of the object of study and we cannot help having some idea of how complicated these mechanisms are. Doing physics is like map-making from a helicopter, you can begin with a bird's-eye view, zoom in later to look at the details; doing psychology is more like making a map on the Lewis and Clarke expedition, right down there in the mud among the trees and the poison ivy.

(2) Experimental Control Is Very Difficult

We cannot breed human subjects like pea plants or treat them like laboratory animals. Moreover, behavior—including mental events—is exquisitely sensitive to countless influences which the chemist or physicist can safely ignore, e.g., whether the experimenter is smiling or sober, male or female, attractive or homely. An old study whose source I have forgotten took advantage of the fact that the same instructor taught two sections of the same course in different rooms. In one classroom, for some reason, there was a faint but distinct odor of peppermint in the air. It was arranged to administer the final examination to half of each class in the peppermint room and half in the room that smelled only of chalk. Those students who were tested in the rooms where they had heard the lectures scored significantly better than their transplanted classmates.

C. Psychology Seeks to Understand the Workings of the Most Complicated Mechanism in the Known Universe

Psychology is the study of behavioral and mental events which, in turn (we assume), are determined by physico-chemical processes in the brain and nervous system. The brain is the most complex mechanism we know of, and its complexity results in large part from the brain's ability to modify itself structurally as the result of learning or experience. The digital computer is a man-made mechanism that shares this remarkable capacity for progressive structural elaboration.

(1) Parametric versus Structural Properties

Both brains and computers are delivered from the factory with a certain standard hardware that is determined by the blueprint, in the case of computers, or by the species plan, in the case of brains. Both mechanisms share the property of almost unlimited structural modifiability.

Entities or mechanisms that have the same structure can be described in terms of the same set of laws. These laws, which we can think of as transfer-functions

or equations relating stimulus input to response output, will contain various constants or parameters. Different systems sharing the same structure can be compared with respect to these parameters, but comparing systems that differ structurally is like comparing apples and oranges. You *can* compare apples and oranges, of course, but you have to know what you are doing and be clear about what you are not doing. We will come back to apples and oranges in a minute.

Computers change or elaborate their structure by being programmed; brains elaborate their structure through experience and learning (sometimes called "programming"). When the structure of a system gets elaborated, so too does the set of laws necessary to describe its functioning. Two Apple computers both running the software called "Lotus 123" are still structurally alike, still can be described in terms of the same laws or the same theory, still can be compared with respect to various parameters. But two computers running different software are to that extent structurally different, march to different laws, and each one will have idiosyncratic characteristics that are not even defined for the other.

The people who study computers and brains have rather parallel divisions of labor. Computers have "hardware experts" while brains have "neuroscientists." The people who write the most sophisticated computer software must have some understanding of the hardware also; they have to understand the laws of the hardware which determine how the structure can be elaborated. They are alike in this respect to some developmental psychologists and to people who study sensation and perception, conditioning and memory and cognitive processing. Finally, the people who use these sophisticated software packages, like *Lotus 123* and *FORTRAN* and *PASCAL*, do not need to know much about the hardware but they must know the rules of the software they are using. Their analogues, I guess, are the personality and clinical and social psychologists. And the big question is, since we have all developed within a broadly similar society, with broadly similar patterns of experience, are we all running roughly similar software packages?

If you use the package called *Word Perfect* for word processing, and *Framework* for spreadsheets, and *PASCAL* for number-crunching, whereas I am using *WordStar* and *Lotus 123* and *FORTRAN*, then our computers may look alike but they won't act alike; you will not understand mine nor I yours. To the extent that our brains are running different programs, no one nomothetic psychological theory is going to be able to account for all of us.

Now, of course, we are always comparing people with one another in a million ways. If we can compare people, sort them out on some dimension, give them each a score, does that not mean that they must be comparable, i.e., structurally isomorphic, i.e., similar systems understandable in terms of the same laws and theory? This brings us back to the apples and oranges. We can compare them in a million ways too—which is heavier or softer or tastes better and so on. When we stop to think about it, many of the most interesting of human psychological

traits are similar to these kinds of comparisons of apples and oranges—I call them “impact traits.”

(2) Impact Traits

An impact trait can be defined only in terms of the impact that the person has on his or her environment, usually the social environment. If you were kidnapped by Martians and studied in their space ship laboratory, they could not assess your Social Dominance or your Interpersonal Attractiveness because those are not so much features of your bodily mechanism or your brain as they are properties of your impact upon other human beings. We can fairly reliably rank-order people for leadership, sales ability, teaching ability, ability to be a good parent—all impact traits—but we do not really expect that the people who get the same rank will achieve that rank in exactly the same way. There are many different ways of being good or bad at each of these things. Just because we can rank people on some dimension does not mean that there is some isomorphic entity or process in each of their brains that determines their score. There are also various ways of achieving any given score on the WAIS. Until it has been shown that g is determined by some unidimensional brain process, the possibility remains open that IQ is an impact trait too.

We can compare apples, oranges, and cabbages using a theory of, say, Produce which contains all the generalities that apply to vegetable foodstuffs. We can compare apples and oranges in terms of a larger set of generalizations which we might call the theory of Fruit. The theory of Apples is richer and more detailed than the theory of Fruit and the theory of MacIntosh Apples is richer yet. The greater the structural similarity between the entities under study, the richer will be the set of generalities that holds true across all members of the set and the more specific will be the predictions we can make about particular entities in the set.

(3) The Nomothetic-Idiographic Issue

It is possible that the general laws of psychology will comprise a relatively small set, that there just are not that many nomologicals that apply across people in general. Perhaps the developmental psychologists will turn out to be better off in this respect; maybe we are most like one another in the ways in which we learn to be qualitatively different from one another. Perhaps the only way to predict individual behavior with any precision or in any detail is idiographically, one individual at a time studied over months or years. To the extent that this is so, perhaps Psychology is really more like History than it is like Biology.

A natural scientist is not embarrassed because he cannot look at a tree and predict which leaves will fall first in the autumn or the exact path of the fall or where the leaf will land. Maybe individual lives are a lot like falling leaves; perhaps there is a very limited amount one can say about the individual case, based

on a knowledge of leaves in general or people in general, without detailed, idiographic study of that particular case and even then it is hard to know how the winds will blow from one day to the next.

Maybe psychology is like statistical mechanics in the sense that we can make confident statements only about the means and variances of measurements on groups of people. We can say pretty confidently, for example, that at least 70% of the variance in IQ is related to genetic variation and that people with IQs of 90 are unlikely to get through medical school. We cannot say that two people with the same IQ must be alike in some part of their brains or that they will achieve comparable success, and we cannot say that a person with an IQ of 140 is going to do something outstanding or useful in the world—that depends on which way the winds blow. We can say that social conservatism, as measured by Tellegen's Traditionalism scale, has most of its stable variance determined by genetic factors. We can say that most of those Americans who favor mandatory testing for AIDS or who admire Oliver North and call the Contras "Freedom Fighters" would get high scores on Traditionalism, but once we start risking individual predictions we get into trouble. Some Traditionalists see the Contras as ordinary mercenaries, Col. North as a troublemaker, and are very nervous about any governmental interference in private lives.

(4) Radical Environmentalism

There are some highly regarded scientists—Leon Kamin, Richard Lewontin, Stephen J. Gould—who believe that our twin research here at Minnesota is immoral, that any findings which seem to indicate that psychological diversity is in any way determined by genetic diversity are either invalid and incompetent or else fraudulent (like the Cyril Burt affair) or both, and that investigators pursuing this sort of research are old-fashioned Social Darwinians at best and probably fascists and racists at worst. These "Anties" have been careful not to assert any specific alternative position that the opposition could criticize; it is easier and safer just to hide in the bushes and snipe at the enemy's breastworks and outposts. If we could capture one of these Anties and put him on the rack and make him say what he really believes, I think it would have to be some sort of Radical Environmentalism doctrine, perhaps along the following lines.

Psychological differences within species of the lower animals are strongly genetic in origin, every dog breeder knows that. A basic postulate of evolutionary theory is that intra-specific variability has been essential to ensure that the species can adapt to environmental change. Behavioral variation has undoubtedly been as important as morphological variability in the evolution of the other mammals. But somewhere in the course of human evolution, probably coincident with the development of human culture, the rules changed. Behavioral variation due to learning and experience began to take the place of variation due to genetic differences until, finally, cultural variation has replaced genetic variation entirely,

in the special case of *homo sapiens*. Unlike dogs or chimps or pigeons, every normal human infant is equipped with a large, almost infinitely plastic brain right off the shelf, all of these brains being made from identical blueprints and specifications. Thus, for our species alone, evolution of the genetic material has achieved a plateau from which the only subsequent evolution will be cultural; phylogeny has ended and ontogeny is all.

If the evolution of the microcomputer continues at its present pace, we might see such a thing happen there. So far it has been useful and adaptive to have available many different sizes and types of computer for use in different applications. New and better designs have made their predecessors rapidly obsolete. One day soon, however, there may come along an Apple or an IBM-PC that is so powerful, so fast, so versatile that hardware development will stop because additional refinements are unnecessary. The only differences then between your computer and mine will reside in the software that we happen to be running.

I think the extreme form of this Radical Environmentalist position is plainly wrong, but there is certainly a large measure of truth in the idea that the proximal cause of much human psychological individuality is learning and experience. If nomothetic theory building requires structural isomorphism within the mechanisms being theorized about—and surely it does, since the point of theory building is to infer what that structure *is*—then the future of personality, clinical, and social psychology depends upon whether the varieties of individual experience produce similar structural elaborations. If our different learning histories yield software packages that differ qualitatively, structurally, from person to person, then perhaps Allport (1961) was right and the core nomothetic theory will be limited to some very general propositions, mostly about learning and development.

Reverting to the computer analogy, there are structural similarities among software programs that might permit a general theory that goes some distance beyond just the structure of the initial hardware. Each of a dozen very different programs may require a subroutine for sorting data, in alphabetical or numerical order, and these sorting subroutines are likely to be quite similar across programs. No doubt there are psychological subroutines which most of us learn and which create reasonably similar structures that will yield reasonably general laws. This would lead to numerous, independent microtheories, each describing software commonalities, held loosely together by a single nomothetic macrotheory concerned with the hardware.

Al Harkness has pointed out to me that many computers come equipped with “read-only” memories or ROMs, innate software packages which serve, among other things, to get the hardware up and running. ROMs enhance the computer-brain analogy by permitting us to talk about innate fears and other instincts, the native ability of the human (but, perhaps, not the chimpanzee) brain to deal with complex linguistic relationships, and the rather extensive pre-programming that

seems to guide child development. Inexperienced goslings show no alarm when a silhouette of a flying goose is passed over head but run fearfully for cover when the same silhouette is passed backward, which makes it resemble a hawk. This implies the innate existence of the same sort of connections or associations that the goslings will later acquire through learning.

In other words, the human brain (and the brains of most "lower" animals) comes equipped not only with hardware capable of elaborate programming but also with certain important aspects of programming already in place. Since we know that there are individual differences in the hardware itself, it seems likely that our ROMs, too, are not always identical, one to another. And it should be emphasized that the brain's ROMs, while perhaps they cannot be erased or written over, can be written around or circumvented. Thus, the incest taboo, which inhibits sexual interest in persons with whom we were reared, whether in a family or in a kibbutz, is not always effective (individual differences) and could doubtless be overcome in most cases if, for some reason, one wished to do so.

(5) Typologies

It is possible that, with respect to personality structure broadly construed, each human individual can usefully be considered to belong to several independent types or taxa and that the laws or theories of these several taxa can be used, alone or in combination, to predict the behavior of that individual in different situations. That is, there may be subroutines (or even ROMs) shared only among the members of a given type.

A type or taxon can be defined as a set of homeomorphic entities. Therefore, a single set of nomologicals, a single theory, will approximately describe all members of a type. For our purposes, it will be useful to modify this definition slightly: we shall define a type as a set of entities that share structural components, i.e., subroutines, that are homeomorphic. Therefore, those aspects of the behavior of these entities which are determined by the structural components that they have in common will be describable in the terms of a single theory. Thus, all radio receivers belong to one type of electronic instrument, all transmitters to another type, and all two-way radios belong to both types. The "kleptomaniac" and the primary psychopath are two quite different subtypes of the weak taxon "criminal."

Since human development begins with a set of homeomorphic entities that differ only parametrically, by what mechanism do people develop structural components shared with other members of the same type? One important insight of modern behavior genetics (one that *would* have impressed Meehl's Ph.D. committee) is that genes influence complex psychological characteristics indirectly, by influencing the kinds of environment the individual experiences or seeks out (e.g., Plomin, DeFries, & Loehlin, 1977; Scarr & McCartney, 1983). The child's temperament and other innate predispositions help to determine how other people

will react to him or her, what sorts of experiences he or she will have, and what imprint these experiences will leave behind. To an important extent—just how important we do not yet know—the brain writes its own software. Since the hardware of the human computer is homeomorphic, since individual differences at the beginning of development are parametric rather than structural then, to the extent that gene-environment covariation is important in development, it is more likely that the structural elaborations wrought by self-selected experience will retain some of that original homeomorphism.

One unique feature of our species is that much of the experience that shapes us is vicarious, derived from stories we hear and probably also stories we make up in our own heads. Much of the primitive person's knowledge of the world comes from stories, traditional myths and experiences related by others. Books and television provide our own children with an almost unlimited range of vivid quasi-experiences which play an important role in shaping their world-view, their knowledge, and probably too their attitudes and personality. Because most of this rich library of vicarious experience is provided cafeteria-style, the opportunity for a modern child's nature to determine its nurture is greatly expanded.

The "cafeteria" metaphor for human experience misleadingly suggests that selections are made stochastically when clearly choices made early on tend to influence choices made later. Because of differences in temperament and native ability, Bill eschews most vicarious experience in favor of active adventure outdoors; Bob is fascinated by science fiction and later by real science; George is addicted to adventure programs; Paul, who is precocious, discovers pornography. What began as mere parametric differences must often lead to real differences in structure. Since human nature is so obviously complicated, perhaps the most we can reasonably hope for is that the varieties of human software packages will be classifiable into a manageable number of homeomorphic types within each of which some rules hold that do not hold across groups or types. (And it is relevant to note that we now have powerful analytic methods for detecting latent taxa in psychometric data, viz., Meehl & Golden, 1982.)

Summary of the Structural vs. Parametric Variation Issue

All sciences have as their objects of study collections of entities or systems, and the job of the science is to describe the behavior of these entities and, ultimately, to develop a theory about the structure of the different types of entities so that their behavior can be deduced from the theory of their structure. This job is relatively easier when the entities are all structurally alike, or when they can be sorted into classes or types within which there is structural similarity. Thus, all atoms of a given isotope of any element are structurally alike; thus, one microtheory fits all exemplars of a given isotope and, moreover, one macrotheory contains the features common to all the microtheories.

The same is true for molecules, the next higher level of organization, although now there are many more types which it is convenient to sort into classes—acids, bases, nucleotides, etc.—and into classes of classes—polypeptides, proteins, etc. And so we can go upward in the hierarchy—organellae, cells, tissues, organs, mammals, primates—seeking to classify these increasingly complex entities into types that share sufficient structural homeomorphism so that a single structural description, a single microtheory, can provide a usefully general and adequate account of all members of the type or class.

The step from neuroanatomy and neurophysiology to psychology, like the step from computer hardware to software, is a very large and different kind of step from any preceding steps lower in the hierarchy. Entities that are extensively modifiable in structure, whose hardware is designed for structural modification or elaboration, are something *sui generis*, without parallel in science or in engineering. Entities in which the hardware helps to write the software are without parallel at all. We can certainly aspire to create reasonably conventional scientific theories about the hardware, about how the brain's structure can be modified. If it turns out to be true that most individuals within a common culture have been modified in reasonably similar ways, or if they can be classified into a manageable number of reasonably homeomorphic types, then we can have at least crude theories—Produce or Fruit theories, perhaps Apple or even MacIntosh theories—about aspects of the elaborated organism, about personality, interests, and intelligence. We must simply keep trying and find out how far we can go.

D. Psychology So Far Has Lacked Good Paradigms

We talked earlier about the long publication lags in social science journals and the suggestion that we countenance this because we are all digging in separate places on the beach, looking for different things; we do not need to know how anyone else is doing—or what they are doing—and we do not fear that anyone else will scoop us because we know that no one else is hunting where we are or for the same thing—i.e., we lack paradigms. In gold mining, a paradigm consists in the discovery of a deposit, a seam, so that people can get to work, all the technicians who know how to dig, to timber a tunnel, to build sluices for treating the ore, and so on.

Heinrich Schliemann was a paradigm maker; he figured out where to dig for the ruins of the ancient city of Troy. Based on his pathfinding, an army of archaeologists could start doing useful work, had whole careers laid out before them. Many good doctoral dissertations were made possible by Schliemann's essential first steps. It is important to understand that just having the tools for research, for digging, is not enough. You can be smart and well trained, bright-eyed and bushy-tailed, but if you do not know where to dig, you may end up in a dry hole or a mud hole. The hot paradigms currently are, of course, in molec-

ular biology. Any psychology graduate student has the option of transferring to one of those areas where one could have almost total certainty of spending one's career doing useful work identifying codon sequences on the ninth chromosome or etc. The paradigms are there, it is just a matter of digging.

Paradigm-makers are few and far between in every science. In psychology there have been a few—Freud, Skinner, Pavlov, and Piaget, to list some important examples—and there have also been some pseudo paradigm-makers or false prophets—Jung, Hull, and Kohler, for example, able and intrepid adventurers who had the bad luck to return empty-handed (and also Freud, Skinner, Pavlov, and Piaget from another point of view, i.e., implicitly or explicitly they claimed too much).

E. Too Many Go into Research Who Do Not Have a True Vocation

(1) Fact: Most Meaningful Research Is Done by an Elite Handful

a) Price's Inverse Square Law. In a 1963 book called *Little Science, Big Science*, Derek de Sola Price pointed out that, going back into the nineteenth century, rates of scientific publication have followed, approximately, an "inverse square law" in the sense that the number, N , of scientists producing k papers is approximately proportional to $1/k^2$. This means that for every 100 authors who produce one paper, 25 will produce two, 11 will write three papers, 6 will write four, and so on (1 of the 100 will manage as many as ten papers). This model suggests that about 50% of all scientific papers are produced by about 10% of the scientists—and we're including as "scientists" not all the graduates or Ph.D.s but only those who have published at least one paper. The modal lifetime number of publications for Ph.D. psychologists is zero.

b) Publication by Psychologists. Out of 20,000 first authors in APA journals over a five-year span, Garvey's APA study found that only 5% appear twice in that five years; less than 2% average one appearance per year—i.e., only about 400 authors publish once per year in APA journals. Using a different data set, George Miller found a similar result, namely that most of the lasting work in psychology was done by a core group of about 400 individuals. Myers (1970) found that half of all authors of articles in psychological journals were cited once or less over the next six years.

(2) The Ortega Hypothesis

Jose Ortega y Gasset, a Spanish philosopher who died in 1955, described the world of science as a kind of beehive:

For it is necessary to insist upon this extraordinary but undeniable fact: experimental science had progressed thanks in great part to the work of men astoundingly mediocre, and even less than mediocre. That is to say,

modern science, the root and symbol of our actual civilization, finds a place for the intellectually commonplace man and allows him to work therein with success. In this way the majority of scientists help the general advance of science while shut up in the narrow cell of their laboratory, like the bee in the cell of its hive, or the turnspit at his wheel. (Cole & Cole, 1972)

In their interesting *Science* paper, the Coles point out that the common view accords with Ortega's, that science is an ant hill or termite colony kind of enterprise, with multitudes of anonymous workers each contributing essential efforts.

Another version from Lord Florey, a past-president of the Royal Society:

Science is rarely advanced by what is known in current jargon as a "breakthrough"; rather does our increasing knowledge depend on the activity of thousands of our colleagues throughout the world who add small points to what will eventually become the splendid picture, much in the same way the Pointillistes built up their extremely beautiful canvasses.

Any large city works, to the extent that it does work, on this principle of the termite colony. So does the world of business and commerce under the free enterprise system. The postulate of free enterprise economists is that this is the only way that the world of commerce can work at all effectively.

Cole & Cole (1972) investigated whether this description actually fits the enterprise of physics by examining the patterns of citations of other researchers in papers published in the physics journals in 1965. They discovered that, at least in 1965, 80% of citations were to the work of just 20% of physicists. They took a "representative" sample of 84 university physicists, got their "best" paper published in 1965, looked at the 385 authors whom these 84 cite. Sixty percent of the cited authors were from the top nine physics departments, 68% had won awards on the order of the Nobel Prize or election to the National Academy of Sciences, 76% were prolific publishers. "Eminent" physicists, as defined by more than 23 citations of their papers in the 1965 *Physical Review*, cited authors who were themselves eminent; they averaged 175 citations per year in *Science Citation Abstracts*. Even non-eminent authors (those with few citations, few publications) cite mainly this same set of eminent authors.

This situation is the same but more so in psychology where less than 20%—perhaps more like 5 or 10%—carry the load.

It may be that modern physics and psychology are nontypical sciences in this respect. I think it could be argued that modern biology, or at least some of its branches, does fit the Ortega model, perhaps not his emphasis on "mediocrity" but at least his idea of the busy beehive. Maybe the paradigm idea is really central here. Theoretical physics in the 1960s was running low on paradigms. The experimentalists were turning up all these strange new particles, showing that the

old theories were inadequate, but no new ideas had surfaced. I remember hearing one of the Minnesota physicists say that he was going into administration because the situation in physics was just too chaotic, everyone milling about, scratching their heads, not knowing which way to go.

I think that the elitism that emerges from de Solla Price's and the Coles' analyses should be tempered a bit this way: only a handful of scientists have whatever it takes to be paradigm-makers, to know where to dig. Many more may be perfectly qualified to do good work, useful work, once a paradigm is available.

(3) Serendipity Is Emergenic

It may be that being a good researcher, in the sense of paradigm maker, is an "emergenic" trait (Lykken, 1982; Li, 1987), the result of a particular configuration of independent traits all of which have to be present or present in a certain degree to yield the result. Having a fine singing voice, for example, is an emergenic trait. Being a great violinist or pianist probably requires high scores on several quasi-independent traits; there are lots of people with a good ear or fast reflexes or deep musical insight or good manual dexterity, but one has to have all of these to play at Carnegie Hall. I would guess that successful paradigm-making may be a multiplicative function of brains \times energy \times creativity \times arrogance \times daring \times ??? and perhaps the relative weighting of the components is different for different fields of study. *Chutzpah* is probably a necessary ingredient in many situations; if you don't sell your ideas, they won't make any waves. Barbara McClintock is a case in point. Her Nobel Prize was awarded for work done many years earlier which had not been noticed because she did not sell it. Someone else realized, retrospectively, that she had really pioneered in a currently hot area and did the selling, belatedly, for her.

In fact, I think that what we call genius is probably emergenic. In the biographies of most people of genius—people like Gauss or Shakespeare or Ramanujan, Mozart or Benjamin Franklin or Mark Twain—it seems apparent, first, that they were innately gifted. We have no idea at all what sort of early experience or training could turn an ordinary lump of clay into people like these. Yet, second, the genius does not run in families. The parents, sibs, or offspring of these supernovae usually do not show similar talents, even allowing for regression to the mean. This might indicate that the qualities of genius comprise a configuration of independent, partially genetic characteristics, all of which must be present to produce the result. The first-degree relatives may have some of the components or more than an average amount of all of them, but, as any poker-player knows, being dealt the Ace, King, Queen, Jack of spades plus the nine of diamonds is qualitatively different from being dealt a royal flush in spades. I don't think you have to be a Gauss to be a paradigm-maker, but I do think that the principle may be similar.

(4) Research Is Over-Valued, Especially in the Academy

There is more pay, more prestige, more freedom, more job security for academics who are successful researchers or who at least manage to publish frequently.

a) Meehl's "*Seven Sacred Cows of Academia*." Among these (regrettably unpublished) fallacious postulates is the proposition that a good university-level teacher must be a productive scholar. The two activities are competitive more than they are complementary. It takes much the same kind of intelligence, thought, reading, and insight—not to mention clock hours—to prepare a good lecture as to write a good article or plan a good experiment. (A really good researcher is likely to be a good teacher only because he/she has these abilities and is the kind of person who won't do something at all without doing it well.) Think about people like Isaac Asimov and Carl Sagan, Walter Munn and Gardner Lindzey, or the late Kenneth MacCorquodale. Munn and Lindzey wrote outstanding textbooks, MacCorquodale was a splendid teacher, Asimov and Sagan have helped millions of people to understand science a little better. All were fine scholars and good communicators, all of them might have made less of a contribution if they had allocated more of their energies trying to do original research.

b) *Teaching and Public Service*. These two avenues through which an academic can justify his or her paycheck are at least as important as research, at least as demanding of very similar abilities. Most research discoveries will be made by someone else if you do not do it; e.g., if Watson and Crick hadn't worked so hard on the double helix of DNA, Linus Pauling would have had it in a few more months. Much useful research is not really very brilliant, the only mystery is why one didn't think of it sooner. Yet it must be said that many bright and knowledgeable people never seem to think of these things or, if they do, don't do anything about it, or can't seem to discriminate between the more- and less-promising ideas that they do have and tend to follow up the wrong ones.

Is it better to turn up even a real nugget of new truth (which many would-be researchers never achieve) or to save a marriage, cure a phobia, teach a really stimulating course, influence legislation by persuasive testimony, plant some important ideas in the minds of students, policy-makers, or laypersons?

Over the past ten years or so, I have spent about one-third of my professional time educating the public about the "lie detector": One does not need specialized knowledge to see that most of the claims of the lie detector industry are nonsense and sheer wishful thinking. Senator Sam Ervin, untrained in psychology, realized at once that the polygraph test is a form of "20th Century witchcraft." Yet most people, including many psychologists, cannot see it until someone points it out. Let's say that I am a Grade B researcher: i.e., nothing wholly trivial or flagrantly wrong, some product that is genuinely useful, nothing really great. If I spend about one-third of my time on polygraph-related matters, that means

one-third less production of Grade B research. In exchange, however, quite a few innocent persons who might have gone to prison because they failed polygraph tests were found innocent, quite a few bad guys who might have escaped prison because they had passed “friendly” polygraph tests are in prison. Where there was virtually no scientific criticism of the lie detector on which legislators, lawyers, and judges could draw, now there is a book and more than 40 articles and editorials and these criticisms have been cited in several state supreme court decisions banning polygraph evidence because of its inaccuracy. Minnesota and Michigan now ban the use of the polygraph on employees; I was the only scientific witness to testify on behalf of both bills. A bill for a similar federal statute was passed by the House of Representatives in 1986, in part because of my testimony, and will likely become law in 1988.

Any Grade B psychologist could have done these things and it demands no great personal sacrifice since it is mostly fun to do; I lay no claim to be either a genius or a saint. The point is that this sort of public service work is more useful and valuable than most Grade B research (and *all* research of Grades C through Z). One suspects that most of you young psychologists would be able to find a way to make a similar *pro bono* use of your abilities and training at some time in your careers. One hopes that more of you will seize the opportunity when it comes along and not be hindered by any silly notion that it is nobler in the mind to publish some dingbat paper instead.

c) Research Has Visibility. One reason research is overvalued is that it gets the glory, its fruits are tangible and public—you can count the books and articles and you know who wrote them. Great teaching or brilliant clinical work goes relatively unrecognized. But we do not have to passively accept this state of affairs. If you think you have a knack for teaching, for example, do not hesitate to cultivate it, work at it, give it everything you’ve got. If your knack develops into a real skill it will be recognized and rewarded, especially if the consumer movement finally reaches the Academy and students start demanding competent teaching. If you shirk developing your teaching skills, however, because you’re too busy writing Grade C papers, then both you and your institution will be the poorer.

III. Some Things We Are Doing Wrong That We Have Only to Stop Doing

Mark Twain once told of an elderly lady, feeling poorly, who consulted her physician. The doctor told her that she could be restored to health if she would give up cussing and drinking whiskey and smoking cigars.

“But, Doctor!”, said the lady, “I don’t do any of those things!” Well,

there you have it. She had neglected her bad habits. She was like a ship floundering at sea with no ballast to throw over-board!

We psychologists are in a much happier position than this lady, for we have an abundance of bad habits. Surrounded by difficulties and complexities, we have invented comforting "Cargo Cult" rituals, adopted scientific fads, substituted pedantry for substance, jargon for common sense, statistical analysis for human judgment. The examples we shall have space for here are only illustrative; our bad habits are legion and every one that we throw overboard will make us feel and function better.

A. Use of Scientific Jargon

When I was serving my time on an NIMH research review committee and was assigned to be primary reviewer for a dumb proposal, I found that it was usually sufficient just to translate the author's proposal into ordinary language. "No, is that really what he plans to do? Why that's dumb!" Graduate students planning dissertation projects could save themselves later grief by following this rule: Using only language understandable by an intelligent layperson, state your hypothesis, the ancillary hypotheses necessary for your experiment to be a test of that hypothesis, and your predictions. If, stripped of jargon, such a prospectus fails to sound sensible and promising, forget it. Many examples of how social scientists, including some of the most eminent, tend to dress up banal ideas in jargon can be found in Andreski's *The Social Sciences as Sorcery*. I take as my moral for this sermon an excellent phrase of Hans Eysenck's: eschew meretricious and obfuscating sesquipedalianism.

Psychologists, and their psychiatric cousins, are susceptible not only to fads of jargon but to fads of methodology, research techniques, experimental designs, even variables chosen less because of their relevance to some important problem than because they are currently in vogue. In the field of psychopathology research, for example, structured interviews and "research diagnostic criteria" are now a *sine qua non* even though they may not be appropriate to one's application. Most current research on the psychopathic personality, for example, defines the target population in terms of *DSM-III*'s category of Anti-Social Personality although (in my opinion, at least) any group thus defined will be hopelessly heterogeneous, excluding some genuine Cleckley psychopaths while including many persons who are not true primary psychopaths at all. The slavish adoption of *DSM-III* classification has purchased an overall increment in diagnostic reliability at the cost of much specific diagnostic validity.

Some scientific rituals are all right in themselves and mischievous only when they are used as a substitute for thoughtful analysis of one's particular problem. The older psychiatric literature contains many meaningless, uncontrolled studies

of various treatment procedures. When it was realized that many patients get better spontaneously, the idea of an untreated control group was invented. Then someone noticed the placebo effect; it became necessary to let the control patients think they were being treated (e.g., with some drug) when they were not. Finally, someone realized that the clinician rating the patient's improvement also could be influenced by knowing who was on the real drug—hence, the “double-blind” design. This simple, sensible approach would not have taken so long to invent if the people then doing psychiatric research had more of the kind of talent that research requires.

Once invented, the double-blind design became ritualized; as long as your study was double-blind, it must be okay. Example: The financier, Dreyfus, after much psychoanalysis and other psychiatric treatment, discovered that the well-known anti-seizure drug, Dilantin, cured his particular problem (Dreyfus, 1981). Dreyfus financed research on Dilantin's applications in psychiatry. Much money was spent giving unselected psychiatric patients Dilantin according to a double-blind design; the results were essentially negative. But who could imagine that any one drug would produce useful effects in all or most patients? Surely the sensible thing to do in this case would be to look for other people with complaints like those Dreyfus had and try the drug on them. Use of a ritualized procedure seems to blind some investigators, depriving them of common sense.

Another common and dangerous fad is the tendency to take up counter-intuitive research findings and then generalize them to the point where they are not only counter-intuitive but false. Perfectly respectable research has demonstrated that honest eyewitnesses are frequently mistaken. Yet, if the witness had a clear view of a woman's face and he identified her as his wife, his testimony has very strong probative value. It has been shown that psychiatric predictions concerning the “dangerousness” of patients or of criminal suspects are frequently in error. Nonetheless, if a twice-convicted rapist, on bail awaiting trial for a third offense, is charged with rape by yet a fourth victim, it is reasonable for the Court, even without psychiatric assistance, to conclude that this individual is dangerous and to refuse bail on the new charge. Common sense tells us that some kinds of identifications are more certain than others, that predictions can be made more confidently in some cases than in others. One of Meehl's classic papers (1957) provides an elegant analysis of this problem. It is the Cargo Cult mentality, when someone cites a “research finding,” which leads us to renounce common sense and embrace foolishness. We should throw it overboard.

B. Over-Reliance on Significance Testing: The Favorite Ritual

Researchers often do not know what they are looking for or what will turn up—but one goal always beckons, namely, a p-value less than .05, since that is what

it takes to get a publication. Pursuit of statistical significance has become the tail that wags the dog.

I once was outside reviewer on a dissertation from a Canadian university, a rather interesting-sounding study of autonomic responses of psychopaths, neurotic offenders, and normals. I found it impossible to determine how the study came out, however, because there were 75 pages of ANOVA tables, 4th order interactions, some of them "significant" and discussed at wearying length. I suggested that the candidate should be passed since he clearly had been taught to do this by his faculty but that perhaps some of the faculty ought to be defrocked.

(1) The Null Hypothesis Is (Almost) Always False

A professor at Northwestern spent most of 1967 flipping a coin 300,000 times, finding 50.2% heads, significant at the .01 level. At about the same time, Meehl and I did our unpublished "Crud Factor" study. We had available from the University's Student Counseling Bureau computerized responses to an "After High School, What?" questionnaire that had been administered to 57,000 Minnesota high-school seniors. We cross-tabulated all possible pairs of 15 categorical variables on this questionnaire and computed Chi-square values. All 105 Chi-squares were significant and 96% of them at p less than 10^{-6} . Thus, we found that a majority of Episcopalians "like school" while only a minority of Lutherans do (52% vs. 45%). Fewer ALC Lutherans than Missouri Synod Lutherans play a musical instrument. Episcopalian high-school students are more likely to be male than is the case for Baptists.

Fourteen of the 18 scales of the *California Psychological Inventory* (CPI) were developed empirically, by selecting items which differentiated various criterion groups (Gough, 1987). There is no general factor that runs through all of these scales or any substantive theory that predicts them all to be interrelated. Yet the mean of the absolute values of the 144 intercorrelations is about .4. In psychology, everything is likely to be related at least a little bit to everything else, for complex and uninteresting reasons. Therefore, any investigator who makes a directional prediction (A is positively correlated with B, Group X has more Z than Group Y does) has a 50:50 chance of confirming it just by gathering enough N—no matter how fatuous or lunatic his/her theory might be (Meehl, 1967).

Bill Oakes (1975) has pointed out that this may not be as serious a problem for genuinely experimental designs in which groups are truly randomly assigned to treatment and control conditions. In correlational designs (e.g., Anxiety vs. Anality) or in comparisons between self-selected groups (e.g., normals vs. schizophrenics), one is asking if one variable is related to some other pre-existing variable and, for psychology, the answer seems always to be "Yes; at least a little bit, although perhaps not for the reason you think." In a true experiment with random assignment, one is asking whether one's experimental treatment affects most of the experimental group with respect to the measured dependent variable

and in the same way, and the answer to that question can be “No.” Oakes cites an Office of Economic Opportunity study in which 13,000 experimental subjects received two hours per day of special instruction in reading and mathematics for one school year. Compared to 10,000 untreated controls, there was no significant difference in the achievement gains over the year.

But difference scores, like these achievement gains, are notoriously unreliable. If the achievement tests had a reliability of .8 and if, say, the one-year retest stability of the scores for the untreated students was about .7, then the reliability of the difference or gain scores could have been on the order of .3. Then 90% of the variance of both distributions of gain scores might be error variance so that even large samples could fail to detect a true difference between them. I think that the only way a psychologist is likely to fail to refute the null hypothesis with really large samples is by using unreliable measures (which, of course, is easy for a psychologist to do!). And if the null hypothesis is always false, then refuting a null hypothesis is a very weak test of a theory and not in itself a justification for publishing a paper.

(2) Statistically Significant Findings Are Frequently Misleading

I once published an article (Lykken, 1968) examining the claim of another author that a “frog response” on the Rorschach test is evidence that the responder unconsciously believes in the “cloacal theory of birth.” That author reasoned that one who believes impregnation occurs *per os* and parturition *per anus* might see frogs on the Rorschach and also be disposed toward eating disorders. A group of patients who had given frog responses were found to have many more references to eating disorders in their charts than a control group of patients without frog responses. The Chi-square was highly significant. We have already seen why we need not feel the least compulsion to accept this theory on the basis of this outcome, but must we not at least admit that an empirical fact has been demonstrated, viz., this connection between frog responding and eating problems?

Remembering that false facts tend to be more mischievous than false theories, let us ask what is the “fact” that this study seems to have demonstrated. The notion of a valid empirical finding is grounded in the idea of replication. Because this author’s result achieved the .01 level of significance, we say that, if this experiment were to be repeated exactly hundreds of times, then we should be willing to bet \$99 to \$1 that the grand mean result will be non-zero and at least in the direction found by the first author. But not even he could repeat the same experiment exactly, not even once. The most we could do, as readers, is to repeat the experiment as the author described it, to follow his *experimental recipe*; I call this process “operational replication.” But neither he nor we know whether he has adequately described all the conditions that pertained in his first study and that influenced the outcome. If our operational replication fails, the most likely explanation will be that his experimental recipe was incomplete. And his original

significance test provides no quantitative estimate of the likelihood that our operational replication will succeed.

If an operational replication is successful, we still cannot be certain that "Rorschach frog responding is associated with eating disorders." Such an empirical generalization leaps far ahead of the facts in hand. These facts are that patients of the type he studied, who give what he calls frog responses when the Rorschach is administered the way he did it, are likely to have an excess of eating disorders, defined as he defined them, listed in the ward notes of the nurses who worked in his hospital. If we are dissatisfied with the limitations of all these particularities, then we do a "constructive replication." In a constructive replication, we deliberately ignore the first author's recipe and focus solely on the generalization in which he and we are interested. We design our own test of that hypothesis, select our own patients, administer the Rorschach as we think it should be given, define "frog responding" and "eating disorders," and assess the latter, in whatever way seems sensible to us. Only by constructive replication can we reasonably hope to compel respect for any claim we make of having demonstrated a generalizable empirical difference or relationship.

A significance test is like a license on a car; you have to have one before you drive to the APA convention, but only an idiot would invest in an old wrecker just because it has a valid license plate. R. A. Fisher himself made a similar point to the British Society for Psychological Research (Fisher, 1929); significance testing may make a finding more intriguing but it takes replication (constructive replication) to make it believable.

(3) Ways of Staying Out of "Significant" Trouble

a) Make Range, Rather Than Merely Directional, Predictions When we test the null hypothesis, that the difference or correlation is actually zero, against the usual weak, directional hypothesis, that the difference or correlation is, say, positive, then even if our theory is quite wrong our chances of refuting the null hypothesis increase with the size of the sample, approaching $p = 0.5$; that is, the bigger and more expensive the experiment, the more likely it is to yield a false result, a seeming but undeserved confirmation of the theory. If our theory were strong enough to make a point prediction (e.g., the correlation is 0.50), then this situation would be happily reversed. The larger our sample and the more precise our measurements, the more stringent would be the test of our theory. Psychological theories may never be able to make point predictions, but at least, like say the cosmologists, we ought to be able to squeeze out of our theories something more than merely the prediction that A and B are positively correlated.

If we took our theories seriously and made the effort, we should be able to make rough estimates of parameters sufficient to say, e.g., that the correlation ought to be greater than .40 but not higher than .80. Then, at least we should be able to claim that the better the experiment the tougher the test of the theory.

Suppose that a very large and careful experiment yields a correlation within the predicted range; what are the odds of this happening even if our theory is wholly false? I know of no general way to quantify this problem beyond saying that the odds are substantially less than the customary value of 50:50. There are no firm and heaven-sent criteria, only informed human judgment applied to the particulars of this case. If the theory does logically lead to the given range prediction, using auxiliary hypotheses that seem reasonably robust, and if the experiment was truly a tough test, then we must respect the theory *a posteriori* more than the frog response result compelled us to respect the theory of cloacal birth.

b) Multiple Corroboration. Any theory worth thinking about should be rich enough to generate more than one testable prediction. If one makes five reasonably independent predictions and they all are confirmed experimentally, one can claim p less than $(0.5)^5$ or less than about 4 chances in 100 of doing that well accidentally.

c) Comparing Alternative Models. As Sir Karl Popper has pointed out, we should not aspire to show that our theory is valid but, rather, that it possesses more ‘‘verisimilitude’’ than any current competitor and therefore deserves interim allegiance until something better comes along. That is, for any theory, if our tests are sufficiently searching and stringent, the theory must ultimately fail. A more constructive approach, therefore, is to apply equally stringent tests to existing alternative models and to focus subsequent research and development on the model or models that fit the data best. This is the approach of modern biometrical genetics (e.g., Jinks & Fulker, 1970; Eaves, 1982) and of structural-modeling specialists (e.g., Bentler & Bonett, 1980; Cudeck & Browne, 1983).

In most areas especially of ‘‘soft’’ psychology, it is rare for a proponent of a theory to give explicit systematic attention to possible alternative explanations of a data set. Showing that one’s theory is compatible with the trend of one’s data is, as we have seen, only weak corroboration for the theory. Showing that our theory fits the data better than all plausible alternative models, on the other hand, is strong corroboration, strong enough in fact to establish our theory squarely in the catbird seat until such time as a new and more plausible competitor is advanced by someone else.

Example: I have proposed that the primary psychopath is the frequent, but not inevitable, product of a typical environmental history imposed upon a child who is at the low end of the normal distribution of genetic fearfulness or harmavoidance (Lykken, 1957, 1984). In a mental maze task where certain errors are specifically punished, we know that psychopaths avoid errors punished by loss of money (quarters) but do not avoid errors punished by a painful shock. That such findings can be predicted from my theory is encouraging, but the fact that they cannot be predicted by rival hypotheses (e.g., the hypoarousal model or the disinhibition model) is considerably more significant.

d) The Multi-Trait, Multi-Method Matrix (Campbell & Fiske, 1959). We know we ought to distrust most alleged measures of particular traits (e.g., “anxiety” tests), and we also know that method variance accounts for much of the common variance in psychological research. Therefore, we can construct a tougher hurdle for our hypothesis by using several measures of each trait and several methods of measurement. We should also include in the matrix measures of other possible traits that might be producing spurious findings. For example, intelligence tends to be correlated with everything so one should make sure that one’s finding that A correlates with B is not just because both A and B are loaded on IQ. The objective is to show that the common factor measured by one’s four measures of X correlates with the common factor measured by the several tests of Y even after the co-variance produced by Z (e.g., IQ) has been removed. Example: One can reasonably wonder whether many of the interesting findings obtained in research on Kohlberg’s (1984) Stages of Moral Development would remain if verbal intelligence had been partialled out in each case.

e) The Two-Phase Experiment and Overlapping Replication. In programmatic research, which is generally the best kind of research for several reasons, we can use the technique of sequential, overlapping replication. Each successive study replicates the most interesting new findings of the previous experiment and also extends them in new directions or tests some new hypotheses. In the initial attack on a new problem, we can use the Two-Phase Experiment. Phase 1 is the discovery phase, the pilot study, in which we find out for ourselves how the land lies. Since we are not trying to impress or convince anyone else, we include only such refinements and controls as we ourselves believe to be necessary to evaluate the hypothesis. If we decide after running three subjects that some aspect of our set-up should be changed, we change it and roll on. If our planned method of analysis of the data yields mostly noise, we feel free to seek a different method that will yield an orderly result. If Phase 1 produces interesting findings and if, in our judgment, we can now design a full-scale experiment that will yield the same findings, then we move on to Phase 2, the proof or verification phase, the elegant experiment designed to convince others (e.g., journal referees) that our findings are valid.

Assuming that our judgment is good, the Phase 2 experiment will always be better designed and more likely to produce useful results because of what we have learned in Phase 1. If Phase 1 does not work out, we will not feel so committed to the project that we will struggle to wring some publishable but unreplicable findings out of it. Muller, Otto, and Benignus (1983) discuss these and other useful strategies in a paper written for psychophysicologists but equally valuable for workers in other research areas.

Reichenbach’s distinction between the *Context of Discovery* (e.g., the pilot study) and the *Context of Verification* (e.g., the Phase 2 study) is a useful one,

especially for psychologists. Since we should be honestly humble about how little we know for sure, it behooves us to be open and relatively loose in the context of discovery. Just as there are few hypotheses than we can claim as proven, so are there relatively few that we can reasonably reject out of hand. Extrasensory perception is a good example. Having worked for years with hundreds of pairs of adult twins, hearing so many anecdotes of apparent telepathic communication between them, which usually occur in moments of stress or crisis, I am inclined to believe in telepathy—as an individual but not as a scientist. That is, I would be happy to invest of my time and the government's money in what I thought was a promising telepathy experiment. But to compensate for this openness in the context of discovery, we must be tough-minded in the context of verification. Since no one has yet succeeded in capturing telepathy in the laboratory, in discovering a paradigm that yields consistent, reproducible results, telepathy remains just an intriguing hypothesis which no one should believe in *qua* scientist.

(4) The Bottom Line

The best single rule may be Feynman's principle of total scientific honesty. Feynman says:

If you're doing an experiment, you should report everything that you think might make it invalid—not only what you think is right about it [but] other causes that might possibly explain your results. . . . Details that could throw doubt on your interpretation must be given if you know them. . . . If you make a theory, for example, you must also put down all the facts that disagree with it . . . you want to make sure, when explaining what it fits, that those things it fits are not just the things that gave you the idea for the theory but that the finished theory makes something else come out right, in addition. (Feynman, 1986)

This is not nearly so easy as it seems since it is natural to become infatuated with one's own ideas, to become an advocate, to be a much gentler critic of one's own work than one is of others'. Many of us are able to tear other people's research limb from limb while we smile upon our own like an indulgent parent. In fact, I think one *should be* protective at first until the toddler at least can stand erect. But before one lets the little devil out into the neighborhood, one must learn to look at it as critically as others will.

Conclusions

In my junior year in college, I was led to change my major from Chemical Engineering to Psychology by the brilliant teaching of Kenneth MacCorquodale and Paul Meehl and by my discovery, in W. T. Heron's course in Learning Theory, that I was already at the cutting edge of development of this slow-blooming

young science. I have never regretted that decision, for there is nothing I would rather have been—that I could have been—than a psychologist. I am a rough carpenter rather than a finisher or cabinetmaker and there is need yet for rough carpentry in Psychology's edifice. This is a field in which there remain many simple yet important ideas waiting to be discovered and that prospect is alluring. I would rather pan for gold dust on my own claim than climb the executive ladder at the *Glitter Mining Company*.

When we exclude those parts of our enterprise that are really neuroscience or genetics or applied statistics, it has to be admitted that psychology is more like political science and economics than it is like the physical or biological sciences and that those colleges which permit undergraduates to "satisfy the science requirement" by taking a few courses in psychology are helping to sustain the scientific illiteracy of the educated segment of society. We can take (rather weak) comfort in the fact that, if our discipline were as mature as physics is, then psychology would probably be recognized as more difficult than physics. It is certainly harder to be a psychological researcher now than it was to do research in physics in Faraday's time.

The brain-computer analogy seems to me to be provocative and genuinely useful, clarifying the relationship among the traditional sub-areas of psychology and illuminating the deep waters of the nomothetic-ideographic problem. It may even be that the new academic Departments of Computer Science will evolve a structure that foreshadows that of future Departments of Psychology.

It is important that we recognize, acknowledge, and root out the Cargo Cult aspects of our enterprise, the scientific rituals and related bad habits by means of which we have sought to emulate the form, but not the substance, of the hard sciences. Some of the most pernicious of these bad habits involve rituals of statistical inference. My prescription would be a limited moratorium on directional tests of significance. From now until the Year 2000, let us say that research reports submitted to psychological journals must include either tests of range, rather than mere directional, predictions or else systematic comparisons of alternative hypotheses. I think these latter, more powerful techniques are potentially within our grasp, but they are new and harder than the nearly futile null hypothesis testing to which we have become addicted. If my idiosyncratic and sometimes overstated critique does nothing else, I hope it illustrates at least that Psychology is truly better situated than Mark Twain's ailing lady who had no bad habits she could jettison in order to regain her health.

References

- Allport, G. W. (1961). *Pattern and growth in personality*. New York: Holt, Rinehart and Winston.
- Andreski, S. (1972). *Social sciences as sorcery*. London: Andre Deutsch.
- Barry, W. M., & Ertl, J. P. (1966). In F. G. Davis (Ed.) *Modern educational developments: Another look*. New York: Educational Records Bureau.

- Bemporad, P. E. (1967). Perceptual disorders in schizophrenia. *American Journal of Psychiatry*, *123*, 971-975.
- Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structure. *Psychological Bulletin*, *88*, 588-606.
- Campbell, D. P., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, *56*, 81-105.
- Chalke, F. C. R., & Ertl, J. P. (1965). Evoked potentials and intelligence. *Life Sciences*, *4*, 1319-1322.
- Cole, J., & Cole, S. (1972). The Ortega hypothesis. *Science*, *178*, 368-375.
- Cudeck, R., & Browne, M. W. (1983). Cross-validation of covariance structures. *Multivariate Behavioral Research*, *18*, 147-167.
- Davis, F. B. (1971). *The measurement of mental capability through evoked potential recordings*. Greenwich, CT: Educational Records Bureau.
- Dreyfus, J. (1981). *A remarkable medicine has been overlooked: With a letter to President Reagan*. New York: Simon & Schuster.
- Eaves, L. J. (1982). The utility of twins. In E. Anderson, W. A. Hauser, J. K. Penry, & C. F. Sing (Eds.), *Genetic basis of the epilepsies* (pp. 249-276). New York: Raven Press.
- Feynman, R. (1986). *Surely you're joking, Mr. Feynman!* New York: Bantam Books.
- Fisher, R. A. (1929). The statistical method in psychical research. *Proceedings of the Society for Psychical Research*, *39*, 189-192.
- Garvey, X., & Griffith, Z. (1963). *Reports of the project on scientific exchange in psychology*. Washington, D. C.: American Psychological Association.
- Gould, S. J. (1978). Morton's ranking of races by cranial capacity. *Science* *200*, 503-509.
- Gough, H. G. (1987). *California Psychological Inventory: Administrator's Guide*. Palo Alto, CA: Consulting Psychologists Press.
- Hearnshaw, L. S. (1979). *Cyril Burt: Psychologist*. London: Hodder & Stoughton.
- Jensen, A. R. (1986). In R. Sternberg & D. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definitions*. Norwood, NJ: Ablex Publishing Corp.
- Jinks, J. L., & Fulker, D. W. (1970). A comparison of the biometrical genetical, MAVA, and classical approaches to the analysis of human behavior. *Psychological Bulletin*, *73*, 311-349.
- Jones, M. B., & Fennel, R. S. (1965). Runway performance in two strains of rats. *Florida Academy of Sciences*, *28*, 289-296.
- Kohlberg, L. (1984). *The psychology of moral development*. New York: Harper & Row.
- Lewontin, R. C., Rose, S., & Kamin, L. J. (1984). *Not in our genes: Biology, ideology, and human nature*. New York: Pantheon.
- Li, C. C. (1987). A genetical model for emergences. *American Journal of Human Genetics*, *41*, 517-523.
- Lindsey, D. (1978). *The scientific publication system in social science*. San Francisco: Jossey-Bass.
- Lykken, D. T. (1957). A study of anxiety in the sociopathic personality. *Journal of Abnormal and Social Psychology*, *55*, 6-10.
- Lykken, D. T. (1968). Statistical significance in psychological research. *Psychological Bulletin*, *70*, 151-159.
- Lykken, D. T. (1982). Research with twins: The concept of emergences. *Psychophysiology*, *19*, 361-373.
- Lykken, D. T. (1984). Psychopathic personality. In R. J. Corsini (Ed.), *Encyclopedia of Psychology*, Vol. 2. New York: Wiley.
- Marston, W. M. (1938). *The lie detector test*. New York: R. R. Smith.
- Meehl, P. E. (1957). When shall we use our heads instead of the formula? *Journal of Consulting Psychology*, *4*, 268-273.

- Meehl, P. E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science*, 34, 103–115.
- Meehl, P. E. (1978). Theoretical risks and tabular risks: Sir Karl, Sir Ronald, and the slow progress of soft psychology. *Journal of Consulting and Clinical Psychology*, 4, 806–834.
- Meehl, P. E., & Golden, R. (1982). Taxometric methods. In P. Kendall & J. Butcher (Eds.), *Handbook of research methods in clinical psychology* (pp. 127–161). New York: Wiley.
- Meehl, P. E., & Rosen, A. (1955). Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychological Bulletin*, 52, 194–216.
- Muller, K. E., Otto, D. A., & Benignus, V. A. (1983). Design and analysis issues and strategies in psychophysiological research. *Psychophysiology*, 20, 212–218.
- Myers, C. R. (1970). Journal citations and scientific eminence in contemporary psychology. *American Psychologist*, 25, 1041–1048.
- Oakes, W. F. (1975). On the alleged falsity of the null hypothesis. *The Psychological Record*, 25, 265–272.
- Plomin, R., DeFries, J. C., & Loehlin, J. C. (1977). Genotype-environmental interactions and correlations in the analysis of human behavior. *Psychological Bulletin*, 84, 309–322.
- Popper, K. R. (1962). *Conjectures and refutations*. New York: Basic Books.
- Rescorla, R. A. (1987). A Pavlovian analysis of goal-directed behavior. *American Psychologist*, 42, 119–129.
- Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype-environmental effects. *Child Development*, 54, 424–435.
- Summers, W. G. (1939). Science can get the confession. *Fordham Law Review*, 8, 334–354.
- Tulving, E., & Madigan, G. (1970). Memory and verbal learning. In *Annual Review of Psychology*, Vol. 29.
- Watson, J. S. (1982). Publication delays in natural and social-behavioral science journals: An indication of the presence or absence of a scientific paradigm? *American Psychologist*, 37, 448–449.
- Weiner, A. S., & Wechsler, I. B. (1958). *Heredity of the blood groups*. New York: Grune & Stratton.