

# Everyday Life as an Intelligence Test: Effects of Intelligence and Intelligence Context

ROBERT A. GORDON  
*Johns Hopkins University*

To show why the importance of intelligence is often misperceived, an analogy between single test items and single nontest actions in everyday life is drawn. Three requirements of good test items are restated, and the analogy is employed to account for underrecognition of the importance of general intelligence in everyday actions, which often fail to meet the requirements and thus fail as intelligence measures for reasons that have little to do with their dependence on intelligence. A new perspective on the role of intelligence in nontest actions is introduced by considering its operation at three levels: that of the individual, that of the near context of the individual, and that of entire populations. Social scientists have misunderstood the operation and impact of IQ in populations by confining attention to the individual level. A population-IQ-outcome model is explained that tests for the pooled effects of intelligence at all three levels on differences between two populations in prevalences of certain outcomes. When the model fits, the difference between two populations in the outcome measured is found commensurate with the difference in their IQ or general intelligence distributions. The model is tested on and found to fit prevalences of juvenile delinquency, adult crime, single parenthood, HIV infection, poverty, belief in conspiracy rumors, and key opinions from polls about the O.J. Simpson trial and the earlier Tawana Brawley case. A deviance principle is extracted from empirical findings to indicate kinds of outcome the model will not fit. Implications for theories of practical and multiple intelligences are discussed. To understand the full policy implications of intelligence, such a fundamentally new perspective as that presented here will be needed.

We all make mistakes in life, and Alexander Pope's "To err is human" is a familiar refrain. There is good reason, however, for supposing that the probabilities of making a mistake in any given situation, independent of experience, vary from individual to individual according to IQ or score on any good test of *g*, the general intelligence factor. This would help explain why "some people make more errors than other people" (Senders & Moray, 1991, p. 69). Full recognition of this probability differential is blunted by the fact that, although life in some ways resembles a test of general intelligence, life departs in many ways from the formal requirements of a well-designed psychometric instrument. Combined with age differences in experience (which can easily be mistaken for differences in intelligence)

---

Direct correspondence to: Robert A. Gordon, Department of Sociology, Johns Hopkins University, Baltimore, MD 21218.

---

INTELLIGENCE 24(1) 203-320  
ISSN: 0160-2896

Copyright © 1997 Ablex Publishing Corporation  
All rights of reproduction in any form reserved

and with age differences in cumulative lifetime risk (which can let the histories of younger and hence less exposed persons seem more error free than those of older, more exposed ones of equal intelligence), such departures from psychometric rigor obscure the role of *g* but do not negate it.

There may well be other measurable human traits that affect the probability of making mistakes, but this article is about general intelligence (*g*). Although the full set of everyday situations in which intelligence comes detectably into play cannot be specified in any simple manner, the article makes a start by laying out relevant considerations and examining illustrative empirical examples.

***Intelligence and the Probability of (Not) Making Errors.*** Differences in probability can be important, even if the probability of an unwelcome outcome never falls to zero at even the highest levels of intelligence. Higher intelligence reduces but does not eliminate the risk of error. Mistakes among the highly intelligent in simple matters, therefore, contrary to some gleeful popular misunderstandings, do not refute the notion that intelligence differences affect probabilities of error.

The probability of error is also a function of the complexity of the task. Task complexity must be held constant when making comparisons between individuals at different levels of intelligence. "Some estimates of error rates have been made and vary from actions that are almost always incorrect to as few as one error per 10,000 opportunities" (Senders & Moray, 1991, p. 62). More difficult tasks are usually tackled by brighter persons, whose errors in some general sense may not be less frequent, consequently, than those of other persons at lower points on the intelligence continuum when everyone is matched to tasks by considerations of competence. It has been estimated, for example, that "as many as 90 percent of industrial and system failures are triggered by human error" (Senders & Moray, 1991, p. ix). However, the rate at which human error contributes to failure may not be much lower at the highest echelons of complex human endeavor; hence, the truth-containing jest that persons rise to their level of incompetence (Peter & Hull, 1969; see also Bugliosi, 1996, pp. 32–35, for an interesting essay, if not an entire book, on incompetence among professionals and other bright individuals, and Dixon, 1976).

Estimates are that fatal iatrogenic injuries to medical patients amount to "the equivalent of three jumbo-jet crashes every 2 days" (Leape, 1994, p. 1851), and that most such injuries are the result of errors, yet medical care providers are among the most highly trained and intelligent of all occupational groupings. The mean IQ of physicians, for example, remained at about 125 across a span of four recent decades (Gordon, 1988). To reduce medical errors further, it has been urged that mistakes be viewed less as moral lapses and character flaws that must be concealed to protect careers, and more as opportunities to look honestly for systemic problems in the organization and conduct of everyday medicine (Leape, 1994).

In that spirit, this article examines the issue of general intelligence in everyday

life, where many tasks tend to be performed widely if not universally. The aim is to elucidate both the practical importance of intelligence within that sphere and some major obstacles to the full recognition of that importance. The effects of such obstacles often exist side by side with a keen, if sometimes reticent, awareness by many individuals of the role of intelligence in its more standard applications, such as schooling and certain occupations.

**Organization of the Article.** The first aim of the article is to show why the role of intelligence in everyday life is often underestimated. The demonstration is conceptual and draws on an analogy that examines outcomes from life as analogs of items within classical test theory. The second demonstration is both empirical and substantive. Its aim is to present a new perspective on how *g* operates within a population, and, using that perspective, to demonstrate a new tool for revealing *g*'s major—and often surprising—lines of influence in social and political life. This effort draws on observations both unsystematic and systematic for concrete evidence of the role of *g* in mundane matters ranging from law-abidingness to opinions about major public controversies. The term *everyday life* will be used to emphasize that the focus is on behavioral contexts in which the role of *g* is documented far less rigorously than has long been the case for such familiar manifestations of *g* as scholastic achievement (Jensen, 1993a) and, more recently, job performance (Gottfredson, 1986b; Hunter & Schmidt, 1982).

Evidence to be presented on the role of *g* in mundane matters will span three levels of analysis: that of the individual, that of the local interpersonal context of the individual, and that of entire populations. To some degree, the full role of *g* at the first and second levels is nested within and cannot always be distinguished clearly from the role of *g* in the more inclusive context. The argument builds toward the context of populations, which serves as a vehicle for new and systematic evidence that ties together and sheds new light on the contexts that precede it. Intelligence research has traditionally focused almost exclusively on the individual level of analysis, but systematic consideration of the second and, especially, third levels of analysis will show that the first level alone is insufficient for comprehending the role of *g* in society and that restricting attention to just the first level can be seriously misleading for policy purposes.

### THE TEST ANALOGY

Mental ability tests are carefully assembled sets of standardized problems, constructed under the guidance of well-developed "test theory." The analogy of a mental test can be applied to life to explore its heuristic value for organizing our thinking about mistakes in everyday matters. Applying the test analogy consists mainly of mapping elements of tests and of classical test theory onto nontest behaviors to determine the extent to which the requirements of good tests and test items are present in the latter. As will become evident, single instances of every-

day behaviors generally do not possess the qualities necessary for revealing the cumulative import of differences in *g*.

Because the test begins with the item, "*the elemental unit of a mental test*" (Jensen, 1986a, p. 302), the test analogy best begins with the item analogy. Many outcomes in life that are studied by sociologists and other social scientists, for example, having been arrested at least once, resemble single-item tests more than they do collections of different items operating together as tests in the customary sense. The item analogy, therefore, is elemental too in the sense that it deals with a unit of analysis that is basic to many fields of study.

### **Three Requirements of Good Intelligence Test Items**

To construct an operational "working definition" of general intelligence based on *g*, the general intelligence factor, Jensen (1986a, 1986b) began at ground level by specifying three properties items must possess to qualify for inclusion in tests of general mental ability. The purpose of such tests, recall, is to distinguish individuals according to their relative level on the underlying trait of intelligence. These requirements contribute to that end. To the degree they are not met, measurement may be less successful and intelligence differences missed, but those differences do not cease to exist just because their effects are less detectable.

First, items must consist of cognitive or mental tasks. Little of the variation in individual performance should be the result of differences in physical capacity. In everyday life, it is not necessary that individual tasks be good measures of intelligence in order to reflect differences in intelligence. As we shall see, all that is required is that they contain a sufficient cognitive component. The purpose of the everyday tasks considered in this article is not, after all, to measure intelligence as well as possible, but to elucidate its systematic *role*, large or small, in nontest behavior.

Second, test items must record performances that can be judged objectively as right or wrong, better or worse, shorter or longer in response time. This requirement accords with the view that human "errors can be defined only in relation to correct and desired behavior," perhaps after "a detailed task analysis" (Senders & Moray, 1991, pp. 59–60). This raises the question, "Defined by whom?" Jensen relied on specially trained observers to help define test item errors, but in everyday life their analog may or may not exist, and conflicting parties may accuse one another of error in addressing the same problem, for example, the question of guilt in the recent O.J. Simpson trial. (In some cases, the accusations may all be correct in that no one has the right answer.) Despite the possibility always of lack of consensus, students of nontest error find reasonable an approach similar to Jensen's: "*If there is general agreement that an actor, Z, should have done other than what Z did, Z has committed an error*" (Senders & Moray, 1991, p. 81).

Defining behavior as right or wrong, or better or worse, is not really an issue in policy realms concerned with crime, HIV infection, or unwed motherhood. Even

the affected individuals themselves often concur with the evaluations of their actions by society broadly: "Humans are generally very proficient in detecting their own errors of action," but "they are much less good at catching their own errors of thinking, decision making, and perception" (Senders & Moray, 1991, p. 78). Disputes over what constitutes a mistake, therefore, are more likely to turn on thinking, decisions, and perceptions, the theoretical provinces especially of intellectuals, than on outcomes of action, a province in which all persons feel more equally at home.

Third, items should be distributed across a range of difficulty appropriate for the target population. Items that are too easy or too hard fail to discriminate among individuals, and hence serve no useful purpose in a test meant to measure differences in capability (although such items may have a place when testing a different population whose average intelligence level is different too). If everyone gets an item right or everyone gets it wrong, in other words, the item has no variance and thus cannot correlate (covary) with either total test score or anything else.

### **Item Requirements and Their Implications Mapped Onto Everyday Life**

*Cognitive Component.* It is often not apparent to persons that cognitive tasks are embedded in many of life's everyday activities, as those activities (say, parenting) often exhibit other, perhaps more salient, facets of content (warmth) that seem to outweigh any cognitive component (judgment) until the latter comes forcibly to attention (as when a warm parent leaves small children unattended with matches in the house). Empirically, however, such noncognitive facets may individually contribute much less than *g* does to variance in overall or long-term task performance (child safety) and, especially, to the task-as-item covariance (i.e., what is common to all such tasks) upon which any test depends for its reliability.

When items all measure something in common, even to a low degree, the total score for many items reflects mostly that common element (cf. Lubinski & Humphreys, 1997). Whatever is specific to each item, no matter how substantial for that particular item, fails to find its match in other items and so fails to accumulate in the total score to the same degree as the common element. What this means is that the cognitive component of a class of tasks (parenting), perhaps even the same task repeated over and over (dealing with child misbehavior), may not be recognized when the task performances are viewed individually, but will emerge from their sum. Human perception, unless schooled by special training, tends to define everyday tasks in their discrete, disaggregated forms, in which the cognitive component is often close to its minimum possible value, rather than as aggregations of similar behaviors.

A simple equation (e.g., Jensen, 1980a, p. 67) based on the aggregation of the variances and covariances of component tasks reveals that one can obtain an im-

pressive measurement of  $g$  from, if not nothing, next to nothing. Hence, the sum of 90 items that correlate with one another only .09, on average, will have a reliability coefficient of .90 (Stanley, 1971, pp. 395–396). In everyday life, performance of a single task with a similar equivalent forms reliability of only .09 is unlikely to attract much notice as a reflection of intelligence, but the item statistics in question happen to apply to the Scholastic Aptitude Test–Verbal of that time. This suggests that if performance in 90 well-chosen everyday tasks were summed in the same manner, a relation to intelligence might become more apparent. Prolonged association with a particular individual can permit the 90 observations to accrue if one is able to recall, correctly label, and synthesize so much data, as can systematic records maintained for research or bureaucratic purposes.

Impediments to recognizing the  $g$  component in single everyday life situations, even those that are clearly mental, have their psychometric parallels in the familiar acknowledgement that even “the very best item is loaded with situation-specificity or error” (Green, 1978, p. 665; Humphreys, 1992), and so a single item “measures intelligence much *less* than it measures a number of other factors” (Jensen, 1980a, p. 128). The correlation of a test item with  $g$  is typically modest, because one can get it right for a variety of reasons unrelated to one’s general intelligence. These would include cognitive or noncognitive traits representing factors other than  $g$  that the item also happens to tap (e.g., spatial visualization ability), blind guessing, special training or experience, fortuitous exposure to key information, and cheating (McCabe & Trevino, 1996), say, by receiving help during the examination from another person. Each of these non- $g$  influences has its potential parallel in the everyday outcomes that are counterparts to test items, but it may not cumulate in aggregate data to the same extent that  $g$  does, except in the case that a good helper is always there as a supplemental source of  $g$ .

In the important special case of help, the helper role might be conceptualized as a sociological variable unrelated to  $g$ , even though, from a wider perspective,  $g$  is very much involved, although now it happens not to be the  $g$  of the individual tested (the proband or focal individual). Specialists in the study of individual differences are apt to regard the helper’s  $g$  as an unwelcome source of interference in predicting a proband’s future rather than as an essential element in the study of  $g$  broadly. Sociologists, on the other hand, typically view a helper’s contribution to explained variance of outcomes as evidence against the importance of  $g$ , as though  $g$  were not involved in the ability to provide good help. All sources of help and hindrance are apt to be subsumed by sociologists under the heading of “social environment,” considered distinct from  $g$  (e.g., Fischer et al., 1996).

Help from another individual, the everyday life counterpart to cheating on a test, but without the stigma of impropriety, is a frequent occurrence, so frequent in fact that Goodnow (1986, p. 86) intimated that the productive use of such help might well be incorporated into the definition of intelligence itself, a move that would impede the study of individual differences as now pursued. Choosing wise-

ly whom to consult requires discernment, however, which may depend on knowing enough about a problem to judge the quality of advice one receives, itself a reflection of competence. "Proficiency at recognizing expertise has important implications for group performance" (Littlepage, Schmidt, Whisler, & Frost, 1995, p. 887), and experimental groups vary in this respect, suggesting that Goodnow may have been responding, in part, to the  $g$  underlying such proficiency, a kind of consumer behavior, which enters into the outcome, so to speak, now through the back door, but in a manner that blunts recognition of the role of  $g$  as commonly conceived.

Reasons unrelated to intelligence for getting an item right can as easily become reasons for getting it wrong, such as chance lack of exposure to key information from associates and acceptance from them of misinformation (mistraining, if the presentation of misinformation is systematically organized). Possibly originating at the topmost rung of formal or informal organizations, errors in recognizing expertise and in failing to reject nonexpertise can thus propagate other errors throughout a social network, adding to non- $g$  components of behavior everywhere.

Underrecognition of the cognitive component is abetted by numerous other features of the phenomenology and organization of intelligence in everyday life. Not least among these is the fact that there certainly are everyday activities in which a cognitive component is thought crucial, but such activities are set somewhat apart by our culture in special categories. The result is that the residual activities may unthinkingly be consigned by laypersons to the noncognitive realm by virtue of the contrast. Education and learning have, of course, been the chief repositories of tasks viewed as essentially cognitive, and it is only in recent times that this view has been broadened among specialists to include performance within jobs and participation in crime.

Although Singaporean mothers were able, when specifically asked, to perceive the role of intelligence in a variety of children's behaviors, such as "Shows common-sense" and "Sizes up a situation badly," it was "Learns quickly," a typically academic marker, that had the highest  $g$  loading in a factor analysis of 55 ranked items (Nevo & Khader, 1995). Between first, third, and sixth grades, academic skills increasingly come to dominate children's attention as examples of what it means to be smart or intelligent (Yussen & Kane, 1985, Tables 2-3). It would appear that social perception of the role of intelligence is drawn toward outcomes with the highest  $g$  loadings, which is not surprising, but it may sometimes be tacitly misconcluded as a result that other outcomes have no  $g$  loadings at all when their loadings are simply not as high. Just as individuals may often be assigned too hastily to only two categories on the  $g$  continuum, say, qualified and unqualified, so may the  $g$ -loadedness of outcomes be falsely dichotomized.

A crucial final point is that, as in aggregate data, repetition of a single task or response by multiple persons can produce regularities in percentages and averages

that are as reliable in life (and as indicative of the operation of  $g$ ) as the results of multiple tasks presented to a single person on tests. If two populations differ in the average  $g$  that they bring to a repeated single task in everyday life, reliable group differences in average performance will emerge, just as group differences emerge in rates of passing an individual test item. Support for the role of intelligence from aggregate data, when uncovered, thus makes it possible to work backward to the inference that intelligence was very likely an influential component of the individual behavior so aggregated if that was not already an accepted view (e.g., Gordon, 1976). This inference will be put to use in connection with a population-IQ-outcome model that is introduced at a later point to examine differences between Blacks and Whites (understood to refer to U.S. populations) in rates of certain important outcomes.

***Right Versus Wrong.*** Examples of failing to recognize the cognitive component in behavior and hence to perceive that responses can be unequivocally right or wrong are readily available from familiar criticisms directed at actual intelligence test items. If such misunderstandings are rife even in connection with items drawn from formal intelligence tests, can they be less common under circumstances in everyday life that are far less formal?

Take, for example, complaints concerning a Year IV-6 Stanford-Binet item, which shows three pairs of sharply contrasting drawings of faces and asks, "Which one is prettier?" The item has been held up to ridicule as an example of "aesthetic comparison," which is what it is unfortunately called (Terman & Merrill, 1960, p. 79), rather than of choice between responses that are cognitive and so capable of being judged better or worse unequivocally (Jensen, 1980a, p. 5). Standards of beauty can vary from culture to culture, it was pointed out, and so right or wrong cannot be settled objectively, unlike such responses as "taller" or "shorter." To such critics, the item lacked, no pun, face validity.

Measurement specialists can compile statistics to determine whether the "prettier" question proved differentially more difficult and less valid for various minority cultures possibly having different standards of beauty, for example, Blacks vis-à-vis Whites (Jensen, 1980a, p. 5). Specialists can also check whether scoring the item in a particular direction led to its correlating positively with other items in all groups. (Such analyses indicate that the aesthetic comparison item is not relatively harder for Blacks, and that it functions as a cognitive item.)

Relatively few specialists would also point out, I suspect, that the aesthetic contrast between the Stanford-Binet drawings seemed great enough to have rendered the issue of taste moot in most cultures, leaving the item to function, not as a test simply of aesthetic preference, but as a picture vocabulary test of the word "prettier" and, especially, of the judgment needed to apply the concept that was intended for 4-year-old children assimilated enough to warrant their being tested in English. One could go on to explain that vocabulary proves to be an especially

good test of *g* because understandings of new concepts are usually acquired by inference from their contexts (Jensen, 1980a, p. 146). (Nowadays, one could also point to extensive research on cross-cultural consistency of facial attractiveness and its biological basis. See Cross & Cross, 1971; Cunningham, Roberts, Barbee, Druen, & Wu, 1995; Grammer & Thornhill, 1994; Jones & Hill, 1993; Langlois et al., 1987.)

Item statistics alone, unfortunately, rarely prove intellectually satisfying to laypersons as evidence of cognitive content and rightness or wrongness, even to those who are mathematically sophisticated (e.g., Houts, 1977), and so statistics need to be coupled, as in this instance, with fuller explanations that only a well-verbalized, empirically founded theory of mental ability can adequately inform.

When item statistics are lacking, as they usually are for everyday tasks, mental test experts are understandably loath to support any single specimen of behavior as a reflection of general intelligence. A tremendous conservatism results, which confines expertise narrowly to the test realm and deprives those interested in everyday behavior of the insights that such expertise might afford.

Arguments against real test items based on the supposed indeterminacy of rightness that are demonstrably faulty, such as those concerning the “pretty” item, do not suddenly become more plausible when directed at specimens of behavior in everyday life just because item statistics are lacking to disprove them. What I am arguing against here, and hope to overcome with data, is a double standard in agnosticism among many test defenders concerning the potential *g*-loadedness of items, depending on whether the items appear on tests or in everyday life.

***Range of Item Difficulties: The Problem of Overeasy and Overhard Items.*** Although again deceptively commonplace to test experts, Jensen’s (1986b, p. 109) third provision, that in order to measure individual differences in a group of people, “item difficulty (i.e., percent ‘failing’ the item) must be greater than 0 and less than 100%,” is of profound significance for understanding why the role of *g* in life tasks tends to be underestimated. Many everyday behaviors, such as operating a car, prove so easy for most persons that they seem not to depend on what the layperson thinks of as intelligence at all, and performing them produces no subjective sense of the effort known as “thinking.” Recall the estimate quoted above that some errors occur as seldom as one in 10,000 opportunities. Many such tasks, of course, were overlearned in childhood, when effort would have been more apparent. Adults who commit inexplicable errors on such tasks are greeted with special epithets, suggestive of no intelligence at all. Perhaps it is the private form of this insecurity, aside from the increased sanctions that may apply, that accounts for the observation, “an inability to admit one has been in the wrong will be greater the more wrong one has been” (Dixon, 1976, p. 166).

Research on elementary cognitive tasks (ECTs), although conducted in the laboratory rather than on everyday tasks, provides especially informative exam-

ples of performances misperceived as making no demand on intelligence. ECTs are often so easy (pressing the button beside the light that goes on) that virtually no one gets them wrong, and participants cannot tell the difference between their own better and poorer performances (Jensen, 1980a, p. 691). Sensitive monitoring of reaction times (defined as the interval, in milliseconds, between the light signal and release of one's finger from a home button) reveals, however, that speed of such performances does vary and is reliably correlated with *g* (Jensen, 1993b). Jensen (1980b, p. 109) remarked that the cognitive demands of one particular ECT "are so extremely simple that it seems almost implausible that the procedure could yield any measurements that would be correlated with IQ." The indefinite linearity of performance with IQ upwards (e.g., Hawk, 1970) appears to apply in the downward direction as well when appropriately measured, to include performance on tasks even as easy as these.

By analogy to test items, everyday behaviors meeting similar specifications of easiness will likewise not be seen as reflecting differences in intelligence. In a famous passage, classic reading for generations of sociology students, anthropologist Ralph Linton (1936, p. 115) stated, "*Most of the business of living can be conducted on a basis of habit, with little need for intelligence and none for special gifts.*" Clearly, Linton might better have said, "little need for high intelligence." His chosen words betray a tendency, strong among intellectuals, to equate high intelligence with intelligence, and lower intelligence with its absence, an example of the too common tendency to transform continua into dichotomies. In Linton's passage, the tendency is exploited, consciously or not, to rule intelligence irrelevant to many pursuits.

Reflecting on his previous benightedness, one of my best undergraduates began a paper, "I always thought that one was either smart or not so smart." Multidimensional scaling indicates "that [undergraduate] subjects appear to conceptualize intelligent behaviors into two distinctive categories, intelligent and unintelligent" (Fitzgerald & Mellor, 1988, p. 153). In its more vulgar forms, this unfortunate tendency divides continuously varying persons into extreme categories such as "smart" and "stupid," "superior" and "inferior," and thus constructs a sort of local egalitarianism within each category. This local egalitarianism has its appeal, but its price is an exaggerated discontinuity between persons in the two categories that makes it all the more difficult to face important social problems relating to variation in intelligence.

The behavioral analogs of items at the other extreme of the difficulty continuum present a somewhat different set of problems for understanding the role of intelligence, but with similar result. Behaviors in everyday life that correspond to extremely difficult items, ones that a given individual or population may have virtually no chance of performing correctly, are widely seen as depending on at least a certain threshold of intelligence, beyond which differences in esoteric forms of learning are usually held to account for individual differences in perfor-

mance short of certified, original genius (e.g., the acclaimed work of an Einstein). Genius itself is often viewed as representing a radical discontinuity in human performance and hence in intelligence (Howe, 1990, pp. 22, 31).

Difficult behaviors other than acts of genius often come packaged together as special forms of expertise in roles that Linton termed "achieved statuses." Typically, such statuses are linked in our society with competition for educational credentials widely recognized as valid markers for above average intelligence. Despite, or perhaps even because of, such markers, laypersons have trouble distinguishing more intelligent from less intelligent professional performances unless the difference in performance is gross. This problem blurs for them the continued importance of the intelligence dimension in its upper reaches even though there is no evidence that performance ever stops being linear with mental ability, and it thus inflates the value of professional credentials as distinct from actual performance (e.g., Bugliosi, 1996).

The relevance of intelligence in difficult matters is further obscured even within the same specialty by the fact that solutions of equal quality may reflect large, but hidden, individual differences in problem-solving time necessary for arriving at them. I am told by a colleague that when a distinguished physicist was asked how he came to outshine much brighter friends, he replied, "It's true. Back in school, they could do whatever I did in half the time. But now I've got the time" (A. Pevsner, personal communication, 1992).

The balkanization of comparisons that follows from intellectual specialization transforms expert behaviors into the analogs of items on narrow, curriculum-specific achievement tests rather than of items on broad intelligence tests. It may prove difficult for several reasons, therefore, to determine from their respective professional labors whether, for example, a given physician is brighter than a given lawyer, or a given physicist brighter than a given anthropologist, special cognitive abilities such as mathematical reasoning aside (Jensen, 1980a, p. 174).

Attempting to make high-end discriminations by resorting instead to the more widely shared challenges of ordinary life, such as cooking, housekeeping, and baby-tending, which in some cultures, according to Linton, are performed by males, is of little help, with the significant exception of spoken vocabulary, because such routine items are rarely difficult for the very bright. This effect may have blinded Linton to the need for intelligence in such activities. In test parlance, mundane life lacks sufficient "top" or "ceiling," that is, lacks items at a sufficiently high level of difficulty to reveal clearly the advantages of high intelligence over average intelligence, unless one applies ingenuity à la Martha Stewart, the popular guru of style in all the details of homemaking, and thus redefines tasks so as to make them more challenging.

***The Import of the Three Considerations.*** Jensen's three familiar criteria acquire new vitality when applied to the behaviors of everyday life, where, often-

times, researchers are interested in the manifestation by many persons of but a single form of behavior at a time, such as criminality or responses to a survey research item. Like many persons, if perhaps not more so, social scientists may be skeptical about the relation between the single behavior and intelligence. Just as skepticism about single items adds up to skepticism about total scores on tests, skepticism about single behaviors eventually adds up to skepticism about the influence of intelligence in everyday life broadly.

Applying the test item analogy, however, relatively discrete behaviors, often based implicitly or explicitly on decisions that vary in quality, can be viewed as corresponding to individual items on a test of mental ability. To judge the intelligence of individuals with high accuracy without testing them (or asking their occupation or schooling, which can serve as rough guides), one must interact with them extensively. Prolonged interaction can be viewed as a way of building up the number of items entering into the judgment. Even for single outcomes, however, we can anticipate stabilities in everyday data akin to those found for real test items when items are repeated by many persons. The high reliability of the rank order of aggregate passing percentages of a set of test items, when many individuals have responded to the same items, is one example of such stability in test data. Correlations of such item difficulties between random halves of a large sample, for example, male and female halves, tend to be about .97 (Miele, 1979). Correlations between groups with different average levels of passing items are often just as high or only slightly lower (e.g., Gordon, 1984, Tables 2, 13, 14).

Like responses to formal test items, the degree of success in performing everyday behaviors can reflect what the psychometrician recognizes as defining attributes of *g*: abstract reasoning, general information and hence the capacity to acquire knowledge, the ability to deal with complexity, judgment, and, finally, the estimation of probabilities and the judicious weighting of conflicting priorities (Edgerton, 1993, p. 222; Gottfredson, 1997; Jensen, 1993b; Snyderman & Rothman, 1988, Table 2.3). In principle, therefore, the cumulative effect of responses to many single situations by one individual can add up to a powerful test of the role of intelligence throughout life, and so can responses to a single situation when aggregated over large numbers of individuals.

### THREE LEVELS OF ANALYSIS

Having described phenomenological impediments—as distinct from politically motivated impediments—to recognizing the role of intelligence in everyday life, I turn now to demonstrations of that role at each of three fundamental levels of analysis: that of the individual, that of the local interpersonal context of the individual, and that of differences between entire populations. The three levels represent distinguishable but interdependent paradigms for examining intelligence in everyday life at successively higher levels of aggregation of persons and effects.

At each stage, the role of intelligence becomes more prominent, with the second level providing a crucial theoretical bridge between the first and third levels. The final product offers a fundamentally new perspective on the role of intelligence in society.

### THE LEVEL OF THE INDIVIDUAL

Almost all research on intelligence has been focused upon the individual level of analysis. For studied outcomes, research usually takes the form of correlating a measure of *g* with the outcome. For several reasons, some made understandable by the previous discussion of the test analogy, the theoretical value of such correlations is often underestimated. First, behaviors are rarely observed at the lowest level of performance, which would make their dependence on intelligence more apparent, and the correlations more convincing, because society is usually structured to prevent such poor performances from occurring. Second, performance failures, when witnessed, are often attributed to superficial causes, for example, not planning ahead, that are formulated in a manner that conceals the role of intelligence behind noncognitive, often motivational, terminology. Third, modest correlations that do get reported between IQ and outcomes are often dismissed as too inconsequential to motivate theory. This major section responds to these three issues, using examples drawn mainly from criminality, a particular interest of the author that can be followed through all three levels of analysis.

#### Analysis of Examples from Everyday Life

Crime for monetary gain may nowadays be the occupation with lowest entrance requirements. Not surprisingly, therefore, criminal behavior that occurs in everyday life often reveals effects of intelligence on task performance with a clarity that employers seldom see or, if they do, seldom report except confidentially, as the very existence of such demonstrations can reflect poorly on a firm or on employees who, unlike criminals, are not otherwise *déclassé* to begin with. The analogy of life as a test might be kept in mind as we consider the following two examples.

With one man waiting outside in a getaway car, two men entered a hotel, inquired about a room, and then one asked the night clerk about getting a job there. He was instructed to leave his name, address, and telephone number.

As the man finished writing the information on a scrap of paper, his partner suddenly realized what was going on and crossed out the name, crumpled up the paper and threw it on the floor.

Then the pair drew guns, scooped up \$52 from the cash register and fled, leaving the paper on the floor. (Associated Press, 1974, p. A1)

The police, not believing at first that anyone would leave his real name, soon had all three under arrest.

Fourteen years later, a man undertaking a similar task passed a holdup note to a bank clerk saying, "Keep clam and no won will get harm [sic]" (Associated Press, 1988, p. C9). In this case, what must be a weak correlation between poor spelling and use of one's own paycheck stub for robbery notepaper surfaced, and he too was easily apprehended.

Both writers employed a familiar *modus operandi*, but planned ahead or anticipated poorly despite high risk and evident motivation in other respects to commit their crimes successfully. Whatever tutelage in crime these young adults had imbibed, it had evidently omitted seemingly pedantic but nevertheless crucial clerical details. The need for such details warns us that a job-training curriculum suitable for these men might seem tedious to others, and, even then, who could tell what other pedestrian information might have been overlooked? The engineering concept *foolproof*, often a goal in design but never attained completely, acquires its significance from our inability to anticipate fully just how unwise persons can sometimes be. The job of supervisor requires constant alertness to the possibility of errors by subordinates.

As an infantry officer I never could get my men to stop kicking expended artillery ordnance whenever we were on terrain that doubled as a firing range. But it was another lieutenant, a gunwise friend from Montana, who blew part of his hand off in the field after idly knocking what proved to be an unexploded World War II round against a tree in Germany. He was a smart college graduate, but his tragic mistake underscores the fact that probability of error as a function of intelligence never reaches zero, even for easy problems.

That simple truth probably drives the effort to discover alternative explanations of such rare events by considering common sense or practical intelligence, as everyday errors like my friend's, seemingly unexplainable by his "academic" intelligence, cry out for understanding. But even on academic tests proper, bright individuals will miss an occasional easy item out of carelessness or whatever that they may get right on another occasion. In the normal course of events some less bright persons, with lower total scores, will get the same item right; therefore, one is inclined to suppose, they must possess something the brighter one was lacking—common sense, perhaps? Clearly, invoking a special form of sense or of practicality fails to give sufficient weight to the fact that, given enough opportunity, low probability events happen.

Witness the infamous Leopold and Loeb murder case of 1924, an early "crime of the century" (Gardner, 1958, p. 13). While serving his sentence of life plus 99 years, the precocious Leopold (1958, pp. 107–108) repeatedly scored 205 or higher on alternate forms of Army Alpha, where the maximum was 212. A dropped pair of Leopold's horn-rimmed glasses, his second mistake, had converted what might have been a perfect gratuitous crime into a perfect case. It was the unimaginably improbable nature of the crime, in view of the backgrounds of the accused, that made it sensational.

Had I been transferred to Germany with my unit, I could have been standing nearby while my friend toyed with the round, foolishly holding my tongue so as not to seem schoolmarmish to a fellow officer, perhaps deferring to his greater outdoor and hunting experience. Fortunately, I escaped having to live with that error. In the first robbery, a somewhat more astute companion detected the error and did attempt to correct it, but even he failed the test, one that would not have presented itself to him but for whom he was with (a brother, apparently). The quality of cognitive help obviously depends on the pool of talent at hand, and even examples of individual behavior often cannot be shorn entirely of their local interpersonal context, where the opportunity exists for a second party to compound the error of the first. But part of that interpersonal component depends on what help individuals judge worthy of accepting as consumers of help.

Analyzing further, we see that the one hotel robber failed to check his own overlearned response to a request for personal information and then failed to anticipate the motivation and ability of others to convert the note identifying him into an invaluable clue. Understanding others is a *g*-loaded task, pursued at the highest levels by PhDs, among others. The robber's companion, although somewhat more sophisticated, failed to anticipate that the crossed-out note could as easily be retrieved from the floor for the very same purpose. The redundancy in their pretexts—asking about a room, then about a job—suggests that their prior planning had been less than perfect.

Appropriate actions as elementary as these, the ECTs of everyday life, give the appearance, we know, of not depending on intelligence at all, and hence such errors have long fueled Freudian speculations about unconscious motives: “wanting to be caught (or to fail).” By excluding individual differences in general intelligence, strong egalitarian assumptions create explanatory vacuums that motives, like esoteric theories of practical intelligence, rush in to fill. Because criminals rarely admit to wanting to be caught—most of their behaviors are inconsistent with that hypothesis—the motives posited that so mysteriously work against their self-interest must be unconscious ones in order to remain in play. It is past time, however, for *The Psychopathology of Everyday Life* (Freud, 1901/1965) to make way for the intellectual pathology of everyday life, where the independent variable has the advantage of being measurable. Mistakes of any sort typically run counter to their makers' self-interest.

It could be argued, with truth, that the robbers were under stress in the situations—“Wouldn't you be?” Although stress may have added to their cognitive burden by assigning, for example, a high priority to haste, its presence does not negate the fact that cognitive tasks of some modest complexity, having clearly right or wrong outcomes, were involved. Techniques of recognizing and managing stress are often themselves cognitively controlled: noting situations likely to induce stress, rehearsing until a difficult task becomes automatic, and assigning an even higher priority to remaining calm.

### **A Concept to Reflect the Cognitive Nature of Planning Distance**

Although some authors employ phrases such as “present-orientation,” “range of one’s time horizon,” and “impulsiveness” (e.g., Wilson & Herrnstein, 1985, p. 169) to account for lapses like those of the robbers, an expression from computer chess, first introduced by information theorist Claude E. Shannon, may be a more appropriate heuristic for a theory based on  $g$ , the ability to deal with complexity.

The number of moves ahead that a chess-playing program can examine before settling on its next move, perhaps by algorithmic weighing of consequences and thus the probabilities of opposing moves, defines its *look ahead* (Zobrist & Carlson, 1973). Being discrete rather than shading into the next move continuously, as is typical in other aspects of life, chess moves are easily countable. In chess, the tree of possible moves at each turn (about 30 on average) forks rapidly into many possible sequences even for a short look ahead of four half-turns ( $30^4 = 810,000$ ). Clearly, the amount of information and thus cognitive complexity involved in foreseeing the future and making decisions under uncertainty is proclaimed better by look ahead than by other terms that, even if they can be linked to intelligence empirically, may seem to refer to noncognitive dispositions such as impulsiveness.

In 1973, computer look aheads were not large numbers, perhaps about four with pruning of obvious bad moves. Knowledgeable selection of possible moves that are promising ones to examine enables chess masters to look ahead about three full turns (full turns include an opponent’s possible moves). A look ahead of just one full turn, with perhaps many fewer than 30 key possibilities to consider, might have postponed the hotel robbers’ visit to jail.

Other determinants of planning distance surely exist, but sheer increase of cognitive complexity with both number of moves and need for knowledge of good heuristics for weighing choices deserves full attention. Some social correlates of planning ahead by individuals that are thought to be of subcultural derivation may simply reflect the mental ability levels of other persons in the local cultural surround of those individuals. Indeed, descriptions of local culture may simply paraphrase the cognitive complexity and abstractiveness created in settings populated by individuals with a particular range of mental ability.

### **An Example of How to Evaluate the Importance of a Low Correlation**

The best-seller *America’s Dumbest Criminals*, which often mistakes sheer haplessness for cognitive errors, equivocates over the actual intelligence of the criminals portrayed so as to evade responsibility for poking fun at “the mentally challenged” and to avoid arousing egalitarians: “Dumb criminals are criminals who *act* dumb . . . instead of using the good sense God gave them” (Butler, Ray, & Gregory, 1995, p. 10). One wishes for less hyperbolic terminology and less dependence on the dichotomization of intelligence. The quotation, however, testi-

fies to the utility, for gaining wide concurrence, of judging the action rather than the actor. Determining the internal state of the actor is best left to independent empirical methods, although it is wrong to deny that there is often some basis for inferring the actor's state from a single act.

Notwithstanding the best-selling authors' cautiousness in emphasizing actions over thinking, evidence indicates that the correlation between official juvenile delinquency and IQ at the individual level is about  $-.25$  in samples that are centered near average IQ (Gordon, 1986, Table 1; Moffitt, Gabrielli, Mednick, & Schulsinger, 1981). This is close in magnitude to correlations between IQ and mothers' reports of their children's adaptive behavior in general, which are slightly lower (Oakland, 1983, Table 1). Sample departures from average IQ in either direction make IQ-specific delinquency rates less variable and so lower the individual-level correlation substantially. Understanding the effect of such departures thus accounts for seeming inconsistencies (Gordon, 1986).

Although  $-.25$  is a relatively modest correlation, as one might expect from the multifaceted aspect of everyday behaviors considered as intelligence items, such a correlation is not unimportant when judged in a wider context. A particular ball-player's probability of hitting a home run in any turn at bat, not far from zero, may appear unimpressive in the full range of probabilities from zero to one, but it may happen to belong to Babe Ruth in his greatest home-run season and thus be quite impressive for judging the potency of that batter among causes of home runs. It is always a mistake to seize on the absolute size of a number for the purpose of dismissing or, for that matter, applauding its importance without deeper reflection as to its real meaning. In the present case, there are the following considerations.

First, observed correlations of delinquency with socioeconomic status (SES) or background, a widely accepted explanation, are not higher than correlations with IQ and are sometimes lower (Gordon, 1986; Hirschi & Hindelang, 1977). As the correlation with IQ rises and falls with sample location along the IQ continuum, so does the one with SES, hence it is necessary to examine both correlations within the same sample when judging the importance for delinquency of either correlation relative to the other, lest a lone low correlation with IQ be compared with numerous higher SES correlations from other studies without proper perspective. Given comparability of conditions affecting the two correlations within studies, if one concedes that the correlation with SES is important, one cannot logically maintain the criticism that the correlation with IQ is too modest to be important. *The Bell Curve* (Herrnstein & Murray, 1994) is built around similar comparisons that often show the effect of intelligence to be greater than that of SES background, as usually measured, on a variety of outcomes.

Second, although the correlation of  $-.25$  is weak for predicting individual outcomes, that is not the only purpose of correlations (Gottfredson, 1997). Correlations of that size have 25% of the utility that a perfect predictor possesses over random selection for choosing individuals, in this case so as to reduce the average

rate of delinquency (e.g., Hunter & Schmidt, 1982; Light & Pillemer, 1984, pp. 154–155). This is not a negligible finding. Sociologists have often considered SES and social class as predictors of averages rather than as predictors of individual outcomes (e.g., Merton, 1957, chap. 4).

Third, the correlation in question is a within-group correlation, to be distinguished from between-group correlations that may show a much stronger relationship between the same variables (Lubinski & Humphreys, 1997). Indeed, in a later section dealing with the population level, we shall see that relatively modest within-race correlations between outcomes and IQ of exactly this sort can exist side by side with virtually perfect between-race correlations. A major point of this article will be that the latter have a far greater impact on society than the within-race correlations, as they may reflect group differences in the lifetime prevalence of delinquents or criminals, expressed as Black/White ratios, that are three, four, or five to one per capita.

### **THE LEVEL OF THE LOCAL INTERPERSONAL CONTEXT OF THE INDIVIDUAL IN EVERYDAY LIFE**

Of the three levels at which *g* can be studied, the second is by far the most neglected and at the same time perhaps the most fascinating. Effects at this level have largely been disguised within social psychology and sociology by attributing them to variables having names that obscure rather than highlight their *g* components. At the same time, the kinds of data most required for unmasking the disguise are almost never generated, because IQ measurements would be needed not only on the probands under investigation (already a rarity in nonpsychometric disciplines) but also on at least some members of each proband's social context, even if that context is a relatively restricted one for purposes of study. Such relational data are expensive to collect in any case, and mental testing adds to costs and time.

Nevertheless, known correlations from research conducted for other purposes, particularly kinship correlations (e.g., Bouchard & McGue, 1981), and various findings on social milieux from stratification studies can often help identify plausible sources of average relational effects attributable to *g* over all probands. Once mindful of such contextual considerations and of the role of *g* in various outcomes, it is possible to read between the lines of many research reports for suggestive evidence of effects resulting from intelligence context. A reader mindful of the effect of *g* on unwed motherhood (Berlin & Sum, 1988, Fig. 9), for example, may recall the .5 correlation in IQ between sibs, and thus the appreciable contextual similarity in that variable, when digesting a report of the effects of a childbearing adolescent sister on the sexual behavior and attitudes of a still younger sister (East, 1996). Part of the effect may be what is sometimes called "spurious," in that it is simply the result of sibling similarity in *g* and consequent similarity in sexual outcomes, but part may be explained by the *g* of the older sister net of the *g* of the younger sister. This latter effect would exemplify that of *g*

in the local interpersonal context, and it does not require that the IQ of the child-bearing older sister be lower than that of the younger. It need only be the case that the risk of adolescent childbearing increases as IQ decreases, and that the average IQ of older sisters in one population, holding younger sister's IQ constant, be lower on average than that of the older sisters in some other population.

The intelligence context of a proband can be a help as well as a hindrance. But, as I shall indicate, there are severe limits on help that prevent equalizing its availability by homogenizing the entire population. Moreover, the contextual benefits of high intelligence are not necessarily always linear within any given setting. Effective leadership in groups, for example, seems to require that the gap in IQ between leader and followers not be too great (Simonton, 1985).

### **Milieu Differences in the Quality of Available Help**

Public opinion researchers have long been aware that there is a stratum of society that reads less, sees less, hears about less, travels less far, participates less in the "thought life" of society, shows less interest in relatively abstract things, and focuses more on the trivial interests of life—all adding up to a "portrait of the underdog" (Knupfer, 1947).

They have also found that information campaigns fail despite strong efforts, as the following classic statement by Hyman and Sheatsley describes:

All persons do not offer equal targets . . . Surveys consistently find that a certain proportion of the population is not familiar with any particular event. Offhand, it might be thought that information concerning that event was not distributed broadly enough to reach them, but that this group would still have an equal chance of exposure to other information. Yet, when the knowledge of this same group is measured with respect to a second event, they tend also to have little information in that area. And similarly, they will have little or no information concerning a third event . . . *there is something about the uninformed which makes them harder to reach, no matter what the level or nature of the information.* (Hyman & Sheatsley, 1947, p. 413)

Students of *g* will recognize that Hyman and Sheatsley were describing what they themselves referred to as *general knowledge*, a widely used measure of intelligence that is called Information on the Wechsler Intelligence Scale for Children and the Wechsler Adult Intelligence Scale–Revised (Wechsler, 1974, 1981), where it is second only to Vocabulary in its *g* loading (Cohen, 1959; Kaufman, 1975). The phrase, "something about the uninformed," is largely a circumlocution for intelligence.

Although Hyman and Sheatsley (1947, Table 6) were uncertain of causal direction, in a national survey the mean scores on their knowledge index (i.e., a *g* measure) of persons at each successive step (0 to 8) of interest in eight foreign affairs topics produced a correlation of .99 between average knowledge and average interest. The authors' uncertainty about causal direction reflects the tendency to view interest as a form of motivation independent of ability. Tetrachoric correlations between interest on one issue and interest on another ranged from .40 to

.82, with a median of .58, leading Hyman and Sheatsley (1947, p. 415) to conclude that "interest in foreign affairs tends to be *generalized*." Similar generalization is shown by  $g$ , of course, and this may help explain the tendency of interests in foreign affairs to become more numerous as  $g$  increased. Both knowledge and interest depend on the level of intelligence to be found at various strata, and the lower the stratum, consequently, the harder it will be to find persons able to serve as good sources of information and advice.

Evidence suggesting a *two-step flow of information* emerged from studies of mass communications (Katz & Lazarsfeld, 1955, chap. 14). Persons designated by others as influential in helping them form opinions tended to have greater exposure to print media, regardless of whether the topic was marketing, fashion, public affairs, or movies, and this held true even when education was controlled (Katz & Lazarsfeld, 1955, Tables 54–55). Print media information was thus transmitted in part via an intermediate step. According to recent data, 27% of adults read a newspaper less than a few times per week, and they tend to be more concentrated in lower SES strata (Barton & Jenkins, 1995, Table 3.6).

The contribution of higher social status to opinion leadership proved important only for public affairs (Katz & Lazarsfeld, 1955, Tables 45, 59). Of the four topics mentioned, public affairs, like foreign affairs in the Hyman and Sheatsley study, is likely to demand the most intelligence if one is to command its content and influence others successfully; media choices make it easy to pursue the other topics at any of several relatively independent levels of cost and complexity. In fact, follow-up surveys of persons named as trustworthy experts on public affairs showed that social status of the experts increased as the interviewers worked their way upwards through successive layers of expertise. The highest socioeconomic status was found among those experts who were named by persons themselves named as experts by persons named as experts by the original sample (Katz & Lazarsfeld, 1955, pp. 283–287; Merton, 1957, p. 412). As it happens, the average intelligence of adults increases with socioeconomic status (e.g., Reynolds, Chastain, Kaufman, & McLean, 1987). The social status findings are thus consistent with other evidence that the intelligence of those regarded as leaders in matters requiring comprehension of complex material increases with the elevation of the intellectual stratum that chooses to follow them (Simonton, 1985). Opinion leadership on the other topics apparently occurred within strata rather than across strata. References to recent studies that link interest in and knowledge about political matters to intelligence and education can be found in Herrnstein and Murray (1994, chap. 12). Correlations between measures of intelligence and various interests appear to vary in size and sign according to the intellectual demands of the interest (Lubinski & Humphreys, 1997, Table 1).

Oversimple though it may be, the two-step model is useful for highlighting the role of the local interpersonal context in facilitating or avoiding mistakes and in determining the contagiousness of error. General descriptions of social relation-

ships provide good reason to expect that the IQ of the individual and the mean IQ of the near social surround tend to be correlated for social structural (Blau, 1960; Kretzschmar, Reinking, Brouwers, van Zessen, & Jager, 1994, p. 575) as well as personal reasons (Baur, 1960; Kandel, 1978; Marsden, 1988; McPherson & Smith-Lovin, 1987; Merton, 1957, pp. 397, 410–411), or, if one prefers, for both macrostructural and microstructural reasons. “[Intelligence] is the key reason I have always felt more comfortable with Jews [a high average IQ group] than with any other ethnic group in America, including my own,” a Black intellectual writes (Patton, 1996, p. 12). Intellectually gifted 11- and 12-year-olds discover unusual camaraderie at summer camps for talented youth, and various individuals report: “I like being with kids who [are] as smart as I am”; “I really feel like I fit in here. We all understand each other because most of the kids are on the same level”; “I love it here because everyone is on the same intellectual level” (Weizel, 1995, pp. 40–41). The incongruous fact is that gifted individuals happily relinquish any advantages they might command in average settings to place themselves among peers who are equally advantaged intellectually. Is this elitism or egalitarianism?

Starting with Sutherland (1939), differential association with lawbreakers has long been viewed as a contributing factor in delinquency, but criminologists have rarely considered that the associations and networks may involve *g* as an operative variable, just as none of the public opinion researchers cited mentioned intelligence, which certainly seemed to be what they were describing. Sutherland (1931), in fact, was influential in opposing any role for intelligence in the explanation of criminal behavior (Herrnstein & Murray, 1994, p. 241).

Indeed, theories of juvenile delinquency (which often strangely omit consideration of adult criminality) are typically scattered under headings such as choice and reasoning, biological, psychological, social disorganizational, lower class based, interpersonal and situational, and self and social control (Shoemaker, 1996). It is not difficult to see that these classifications, which are usually viewed as competing among themselves, could all be different anatomical locations for patting the elephant of *g*—individual *g* as well as contextual *g*. If combined into what criminologists would call a multiple-factor model (Sampson & Laub, 1993, p. 36) or others a multifactorial model (Bailey, 1989), sociological variables representing the seven theory types would probably augment the prediction of delinquent behavior, but *g*, if it were included at all, would be credited only with the contributions of probands' IQs.

Edgerton (1993) found that help from benefactors, sometimes a spouse, was the critical determinant of how well once-institutionalized retarded adults fared in the community. It is safe to assume that the IQs of the benefactors were higher than those of the helped individuals. Recently, the American Association on Mental Retardation abandoned distinctions based on levels of mental retardation, and instead substituted a classification scheme for retarded individuals based on duration and intensity of the support they required from others for successful adaptation: intermittent, limited, extensive, and pervasive (Gresham, MacMillan, &

Siperstein, 1995). Such measures of support received, compared with measures of support given, both perhaps combined into a “net helpfulness index,” could realistically be applied to individuals throughout the IQ range. Destructive as well as constructive support is always a possibility, as when companions facilitate commission of a crime. Public school education without tracking is essentially an attempt to integrate children of various IQ levels into social strata where the IQ level is high.

Although driving a car, as was mentioned, may superficially appear to have little relation to intelligence, public health investigators of the form of mistake known as motor vehicle accidents have found “low rate of intelligence” and “poor judgment” (Iskrant & Joliet, 1968, p. 46) of drivers to be among the causal factors implicated. O’Toole (1990) reported mortality from motor vehicle accidents to be 2.85 times greater among males under IQ 85 than among males over IQ 100. Investigators have had to consider the further likely contribution of social context, the important point for the moment. Iskrant and Joliet couched the contextual contributions, conventionally enough, in terms of visible attitudes and beliefs. Nevertheless, their description indicates how the *g* level of one’s immediate social context, when manifested in “attitudes and beliefs,” can affect the behavior of individuals net of their own *g*. The public health investigators stated:

Every “group” to which an individual belongs has some influence in developing and modifying his attitudes and beliefs. Interpersonal relations (those between parent and child, for example) . . . are influential in determining accident experience. . . . It seems likely that the [bad] example set by parents in the motor vehicle strongly influences the child’s attitude, especially in later life when he himself drives . . . This problem is magnified if a child’s peers, who in turn are influenced by *their* parents’ attitudes, exert similar influences on him. (Iskrant & Joliet, 1968, pp. 42–43)

Clearly, if beliefs and attitudes, not to mention quality of help, can reflect intelligence, then the intelligence of the social surround, which is populated by major categories of actors such as parents and peers, can be an important component of the probability of mistakes and of successful adjustment at the individual level across the IQ continuum. The intelligence context, although never infallible, can often weed out mistaken attitudes and poor forms of help with varying degrees of success, depending on its sophistication. The result is often what is called culture. Data will be presented at a later point, much of it concerning beliefs and attitudes, that can be better understood only by thinking in such terms.

### **Help and the Problem of Reciprocity: The Emergence of Nascent Structure**

The everyday importance of adequate help made evident by the hotel robbers and by considerations of context, already discussed, becomes more striking when we consider that many of the influences on outcomes studied by sociologists, such as parental background, peer networks, and formal schooling, would fall easily under the heading of “help”; elusive forms of help, of what Kropotkin (1902/1989)

called mutual aid, may, in fact, account for at least some of the unexplained variance in numerous outcomes. Almost all of the policy recommendations offered by social scientists could be summarized in just two words: more help. Rare policy exceptions are apt to carry warning labels, such as “tough love,” “bad-tasting medicine,” “self-help,” and “weaning.”

There are, however, limits on normal help. Karl Marx’s famous slogan, “From each according to his ability, to each according to his need!” (McLellan, 1978, p. 75), was one of the most sweeping formulations ever of help as a policy, but it remained an ideal that never was implemented. Helping is supported by social norms (“It is better to give than to receive”), but help as a permanent solution to stubborn differences in intelligence of any size eventually runs up against what Gouldner (1960, p. 171) called “the norm of reciprocity,” which he described as one of the principal components universally present in moral codes. Reciprocity “serves a group *stabilizing* function” (p. 176) and furnishes “one among many starting mechanisms” (p. 177) for social systems (“Got a match?”); its absence, therefore, can have important consequences. The norm of reciprocity obliges one to give roughly comparable benefits to those from whom benefits were received, and its clear violation risks being seen as exploitation. According to evolutionary theorists, reciprocity is demanded especially between genetically unrelated individuals (Irons, 1996).

Although Gouldner (1960, p. 178) allowed for relaxation of the norm in relations with children, the elderly, and “those who are mentally or physically handicapped,” he failed to consider that differences in intelligence throughout the remaining range can often be large enough to pose formidable barriers to fair reciprocation between randomly paired individuals. Society solves this problem in part through the market system of economic exchange and its unequal remunerations, and in part by abhorring random pairing and thus creating relatively homogeneous substructures encapsulated within a diffuse sense of community and nation. Hierarchically arranged substructures, in particular, limit exposure to demands for help that can never be reciprocated, but simultaneously they also limit the quality of cognitive help readily available within structures low in the hierarchy. “By defining the group with which an individual may have intimate clique relationships, our social class system narrows his training environment” (Davis, 1943, p. 609). Such an observation would extend beyond training to all forms of help in everyday life.

Hierarchical rankings often coincide with spatial living arrangements. Maller (1933, Table 2) obtained the IQs of all White fifth-grade pupils in New York City in 1930, and averaged them within each of 273 city Health Areas. The mean IQs of Health Areas ranged from 74 to 118, with *SD* 8.3. Maller (p. 121) viewed the areas as constituting “small communities.” Clearly, the urban community settings varied enormously in their mean levels of intelligence and intellectual resources, but their relative homogeneity compared with random dispersal of persons must have greatly facilitated local reciprocity.

Normative tensions over help, both the need for it and the need to escape demands for it felt to be exorbitant, are more likely to build as the range of

intelligence differences confronted by a society increases. Many people may not realize what the robbery examples only hinted at, namely, that there may be very little in the way of general cognitive problem solving (which excludes problems solvable through special training not available to all) that members of low-scoring groups can achieve that high-scoring groups cannot (e.g., see Lubinski & Humphreys, 1997, Table 7). The asymmetry of this relation sets the stage for violations of the norm of reciprocity whenever high and low groups associate intimately.

Consider an experiment by Laughlin and Johnson (1966) that induced the local equivalent of random pairing. College students were administered the high level Terman Concept Mastery Test (Terman, 1956) and then assigned to low, medium, and high ability groups according to their scores. Students were paired up systematically to represent every combination of the three ability levels and readministered the same test working together, but some members of each level were left to repeat the test alone.

Working with lower ability partners led generally to score improvements, but the more able the partner the greater was the gain. For present purposes, the most telling finding was that medium and high ability individuals improved slightly more working alone the second time than working with low ability partners. As these were college students, and the test was a difficult one, the low group would have scored well above many persons in the general population. Clearly, the medium and high ability groups already possessed virtually all the information and problem-solving ability that members of the low group had to contribute to the task. Among members of all three groups, reciprocity in needed cognitive help in most everyday matters would be virtually impossible to achieve. Only highly particular but relatively easy items of information could serve in reciprocal exchange (e.g., "Do you know what time it is?").

Unfortunately, Laughlin and Johnson did not report how the students felt about working with their partners. Schofield (1982), however, has described classroom behavior of children in a middle school that illustrates relations among help, reciprocity, and ability. Pupils preferred to exchange help with partners of similar achievement level, with whom they could reciprocate and who could reciprocate with them (p. 87). Poor students had little to offer as the brighter ones already knew the answers to easy problems, and so the poorer ones would often be ignored when seeking help from better students (p. 88). This left poor students to turn for help to one another (where they might well imbibe misinformation uncritically). When help was offered to poor students by altruistic better students, the former often rejected it because they were embarrassed not to be able ever to reciprocate (pp. 89–90). Bright students of all races resented attempts to copy in test situations (p. 91), as though feeling exploited. As a result of such tendencies and a large achievement gap between Black and White pupils, helping relationships tended to divide along racial lines (pp. 85, 92), although everyone cooperated well enough with one another in matters unrelated to coursework (p. 86).

Exchanges of direct, explicit help, not to mention true collaboration, are most likely to thrive between individuals unequal in  $g$  when they are separated by only

small steps of IQ, so that gaps are not too big to permit adequate reciprocity. “I didn’t believe in working with anyone who wasn’t my equal,” a senior lawyer states (Toobin, 1996, p. 76). The constraint would impose a positive correlation on individuals between their own intelligence and the intelligence of those from whom they are most likely to receive help on a regular basis, even if some correlation were not already present for social structural reasons. The constraint may, for example, help explain assortative mating for IQ, the average correlation of about .36–.43 between spouses found when many studies are reviewed (e.g., Bouchard & McGue, 1981; Jensen, 1981, pp. 116–118). The upper limit would leave spouses not much different from sibs in IQ similarity and mean separation (overlooking a small mating bias toward higher IQ in husbands). General intellectual reciprocity between spouses would thus occur, on average, over a gap of only about 13 IQ points, but role specialization based on training and interests would provide other opportunities for fair intellectual exchange.

Microstructure, as we have just seen with reciprocity, often contains the seeds of macrostructure (probably because it remains closer to human nature). But macrostructure is the true habitat of entire populations, which can be studied only piecemeal at lower levels. The following sections present a way of studying the effects of *g* in entire populations that merges individual, contextual, and population levels—microscopic and macroscopic phenomena—within a single perspective. The method uses as data differences between populations in rates of certain social pathologies.

### **THE LEVEL OF ENTIRE POPULATIONS—BEHAVIORAL OUTCOMES**

A major concern of research, politics, and social policy in the U.S. is why the Black and White populations differ in rates of good and bad outcomes. Such differences are often attributed, sometimes rather freely, to differences in poverty and to racial discrimination. In this section, a model will be introduced, explained, and tested repeatedly that enables us to consider whether IQ differences often play a strong role in differential outcomes. This model—called the population–IQ–outcome model—incorporates the individual and local contextual levels already described, but expands the latter so as to include all within-population influences that contribute to between-population differences, however remote their source from any proband and however weak the influence may be, and all instead of just selected parts of a racial population.

The population level makes it possible to employ population IQ parameters (mean and standard deviation) in the model in question, which is not possible when dealing with myriad local contexts, each of which may constitute a highly unrepresentative selection from the whole racial population in which it is embedded. Details of contextual effects that can be captured in principle at the second level of analysis are lost to view in the population–IQ–outcome model. In exchange, however, all possible contextual effects attributable to *g* are included and tested

even though they are not directly measured and even though their sources can be conceptualized at best only by classifying them according to major category.

Significant between-group differences in outcomes of the sort to be examined may in the long run loom far more important than individual differences in matters of social policy and of intergroup relations, try as we might to confine attention to the individual level regarding certain issues (Gordon, 1986, pp. 112–114) and to treat persons as individuals rather than as members of groups. It is imperative, therefore, to understand the implications and the magnitudes of group differences in order to address properly the special problems that they pose.

Crime is one example of an outcome, because of the high stakes, in which group differences more than individual differences may affect *de facto* policy, even though *de jure* distinctions are routinely made between (criminal and noncriminal) individuals. Taxicab drivers of both races, to provide but one familiar instance of group-based decisions, often pass up Black males out of reality-based fear (D'Souza, 1995, pp. 250–252; Kennedy, 1994; Roth, 1994, p. 69; West, 1994, p. xv; on predicting dangerousness and reactions to variation in ambient risk levels, see Gordon, 1977, 1982). To give another example, almost everyone is more apprehensive toward strange males than toward strange females.

Public opinion on many social issues is another such outcome, because it exists almost entirely on the aggregate or group level. Individual opinions are often left unsolicited, unspoken, or unreported on a variety of topics, except between close acquaintances. Public opinion polls, therefore, tend to become the voice of group differences—W.H. Auden called them the “voice without a face” (quoted in Crossen, 1994, p. 128)—even though opinion polls invariably document the existence of minority factions within any demographic group. Through the medium of polls, groups address one another.

The behavioral outcomes to be examined are all important ones. At a later point, public opinion on issues of special relevance to race relations, another kind of data entirely, is examined too. Sufficient background is provided to enable contrasts with the usual way of looking at such outcomes to be drawn, and to promote awareness of the far-reaching implications for everyday life of intelligence differences at the population level. Finally, questions concerning the nature of the variable assumed to be intelligence are anticipated and addressed.

### **Various Social Pathologies—The Behavioral Data to Be Examined**

Law-abidingness, family soundness, health, and prosperity are obviously important aspects of everyday life, as well as matters of great public concern. Table 1 displays rates (technically, proportions expressed as percentages) of failure to reach such desired behavioral outcomes in different demographic groups. The rates, as they are called in common speech, consist either of race-specific lifetime prevalences, point prevalences, or approximations to such measures for various outcomes. These statistics are known widely as crime rates, poverty rates, and so on, depending on the outcome in question.

Put simply, lifetime prevalence (prevalence, for short) of an outcome by a given age can be defined as the proportion of individuals who qualify as ever having experienced that outcome within the age range specified (Gordon & Gleser, 1974; Kleinbaum, Kupper, & Morgenstern, 1982). Point prevalence is the proportion of a specified population who qualify (by, for example, being under control of the justice system) at a particular point in time,  $t$  (Kleinbaum et al., 1982). An individual's contribution to lifetime prevalence is irrevocable, whereas an individual who contributes to point prevalence at time  $t$  may not at  $t + 1$  or  $t - 1$ . Prevalence statistics have many important uses, but their greatest value for present purposes is in assessing "the relevance of various theories of the causes of a problem" (Greene, 1974, p. 1).

Listed are prevalences for numerous criteria of juvenile delinquency and adult criminality, and for quite different outcomes, single-parent families maintained by mothers, HIV infection within two age ranges, and two types of poverty relating to children and to their families. All of the foregoing are further specified by race and often sex. The original data summarized in Table 1 were generated over a 50-year period and were based on either large representative samples or what amounts to entire universes of cases.

The various percentages in Table 1 are not confined to any particular level of magnitude but instead are spread throughout the range from .05 to almost 100. Percentage differences in the race-specific prevalences are also dispersed widely, ranging from .57 to 48.0. In brief, Table 1 contains quite heterogeneous data, both as to content and as to magnitude.

**TABLE 1**  
**Race-Specific Prevalence of Delinquency and Crime**  
**According to Various Justice System Criteria, of Single Motherhood,**  
**HIV Infection, and Poverty (%)**

	Blacks	Whites	Difference
<b>Appearance in Juvenile Court, Philadelphia, 1949-50 by age 18.0<sup>a</sup></b>			
1. Boys	50.86	17.86	33.00
2. Girls	15.82	3.35	12.47
<b>Incarceration in a Training School, U.S. ca. 1964, by age 18.0<sup>b</sup></b>			
3. Boys	4.00	1.01	2.99
4. Girls	.82	.23	.59
<b>Philadelphia arrests, Uniform Crime Report (UCR) index offenses</b> <b>(homicide, rape, robbery, aggravated assault, burglary, larceny, and auto theft),</b> <b>1945 and 1958 birth cohorts<sup>c</sup></b>			
Birth cohort I, 1945			
5. Males, by age 18	26.8	8.2	18.60
Birth cohort II, 1958			
6. Males, by age 17	26.0	8.9	17.10
7. Females, by age 17	6.0	1.8	4.20

(continued)

**TABLE 1**  
**Continued**

	Blacks	Whites	Difference
<b>Full lifetime arrest, males, nationwide, before 1967, for nontraffic and UCR index offenses<sup>d</sup></b>			
8. All nontraffic	68.3	46.9	21.4
9. UCR index offense	52.0	14.6	37.4
<b>Males aged 26–33, mean age 29, incarcerated when interviewed, NLSY 1979 through 1990<sup>e</sup></b>			
10. NLSY adult males	13.1	2.4	10.7
<b>Males aged 20–29, under criminal justice supervision (prison, jail, parole, or probation) on a given day in 1989 and 1994 according to the Sentencing Project<sup>f</sup></b>			
11. 1989	23.0	6.2	16.8
12. 1994	30.2	6.7	23.5
<b>Point prevalence of one-parent families with children under 18, maintained by mother, of all families with such children<sup>g</sup></b>			
13. 1970	33.0	8.9	24.1
14. 1980	48.7	15.1	33.6
15. 1990	56.2	18.8	37.4
16. 1993	58.4	20.2	38.2
<b>Prevalence of HIV Infection, January 1, 1993</b>			
Ages 27–39 <sup>h</sup>			
17. Males	3.0303	.7194	2.31
18. Females	1.0204	.0600	.96
19. Both sexes	2.0254	.3897	1.64
Ages 18–59 <sup>i</sup>			
20. Males	2.29	.49	1.80
21. Females	.74	.05	.69
22. Both sexes	1.52	.27	1.25
<b>Prevalence of children living in poverty, of all children, by specific years, U.S. Census or Current Population Survey (CPS)<sup>j</sup></b>			
Official poverty			
23. 1939	96.7	69.4	27.3
24. 1949	87.7	42.2	45.5
25. 1959	65.2	20.2	45.0
26. 1969	44.8	11.6	33.2
27. 1979	37.6	12.5	25.1
28. 1979 (CPS)	40.8	11.9	28.9
29. 1988 (CPS)	43.8	15.3	28.5
Relative poverty			
30. 1939	74.3	34.5	39.8
31. 1949	67.0	21.5	45.5
32. 1959	63.3	18.9	44.4
33. 1969	55.8	17.5	38.3
34. 1979	49.6	19.2	30.4
35. 1979 (CPS)	53.2	17.9	35.3
36. 1988 (CPS)	52.6	22.3	30.3

(continued)

**TABLE 1**  
**Continued**

	Blacks	Whites	Difference
<b>Derived prevalence of families with children living in poverty, of all families with children, 1939 to 1988<sup>k</sup></b>			
Official poverty			
37. 1939	95.4	62.4	33.0
38. 1949	83.1	35.7	47.4
39. 1959	57.0	16.9	40.1
40. 1969	38.5	9.9	28.6
41. 1979	34.1	11.1	23.0
42. 1979 (CPS)	36.4	10.4	26.0
43. 1988 (CPS)	39.9	13.5	26.4
Relative poverty			
44. 1939	69.1	28.1	41.0
45. 1949	61.0	17.8	43.2
46. 1959	55.2	15.6	39.6
47. 1969	49.2	14.8	34.4
48. 1979	45.5	17.1	28.4
49. 1979 (CPS)	48.3	15.7	32.6
50. 1988 (CPS)	48.5	20.0	28.5

<sup>a</sup>From Gordon and Gleser (1974).

<sup>b</sup>From Gordon (1973) as slightly refined in Gordon (1986, p. 122).

<sup>c</sup>From Visher and Roth (1986, Table A-1), presenting data from Tracy et al. (1985).

<sup>d</sup>From Visher and Roth (1986, Table A-6), presenting data from Christenson (1967).

<sup>e</sup>National Longitudinal Survey of Youth data, from Herrnstein and Murray (1994, p. 339), who controlled for age differences within the sample. Their statistic is not quite equivalent to lifetime prevalence of incarceration at least once, because annual interviews could have missed incarcerations briefer than 1 year.

<sup>f</sup>From Mauer and Huling (1995, Table 3); a point prevalence.

<sup>g</sup>From U.S. Bureau of the Census (1994, Table F, p. xiv).

<sup>h</sup>Associated Press (1995), verified by P.S. Rosenberg (personal communication, February 12, 1996).

<sup>i</sup>From Rosenberg (1995, Table 1); essentially similar rates have been reported by Karon et al. (1996, Table 2).

<sup>j</sup>From Hernandez (1993, Tables 9.4 and A-9.4), who adjusted 1939 rates slightly to make them more comparable with rates for other years by taking welfare income into account (p. 238, and Tables 9.3 and A-9.3). Deriving family rates from the adjusted data for 1939 would introduce additional complications. For consistency with derived rates, therefore, results are based throughout on unadjusted 1939 data; the adjusted rates for 1939 yield even better fits for children than the unadjusted rates.

<sup>k</sup>Derived from children's rates and data in Hernandez (1993, Tables 2.4, 9.4, A-9.4), as described in text.

### The Population-IQ-Outcome Model

A simple, but interesting, model can be fitted to the prevalence data of Table 1 that will enable us to assess the extent to which differences in IQ distributions of populations can account for differences in the prevalences. It represents a new way of exploring the joint impact on behavior of the focal individuals' (probands') own IQs and of the IQ levels of the other persons who constitute the probands'

effective social contexts, both proximal and distal. The model is fundamentally different from the usual individual-level models in the study of intelligence. Its objective is to examine the nature of population differences, not individual differences. The two kinds of difference can have different sources and consequences in social life, as should become clear. For instance, the argument that racial discrimination accounts for racial differences in outcomes is a familiar model that proposes a cause that operates mainly between racial populations, rather than within them.

It was, in fact, the discovery through a form of meta-analysis that age–sex–criterion-specific prevalences of White delinquency were remarkably invariant over most of the urban–rural continuum, over long periods of time, and across different jurisdictions that helped stimulate the hypothesis that a relatively invariant cause such as the IQ distribution was involved (Gordon, 1976; Laub, 1983). Such stabilities in rates are more consistent with causes that are relatively unchanging properties of the populations concerned than of the more variable spatial and temporal locales. One such stable property of a population is its IQ distribution, which suggested examining rate differences between populations with different IQ distributions. The first four sex-specific comparisons of Blacks and Whites did, indeed, implicate IQ as a major explanation of the three- to four-fold ratio of Black to White juvenile delinquency rates, and the remaining data in Table 1 enable us to extend those early tests of the model to more observations and to different outcomes.

If each prevalence in Table 1 is treated as an ordinary percentile of a normal distribution, so that a delinquency rate of 50.86% is regarded as the 50.86th percentile, for example, the IQ associated with the percentile in its corresponding population can be determined easily by a simple procedure known to students of elementary statistics. Taking account of the mean and standard deviation of IQ for each race (the IQ parameters that describe the IQ distribution of each), one consults a table of the unit normal distribution to determine with a few simple calculations the IQ associated with that percentile for each race (the simple equation appears in Gordon, 1987, p. 44). The IQ so determined defines the hypothetical critical IQ for that outcome in that demographic group. Hence, the general term, *population–IQ–outcome model*, where populations are understood to be ones with known differences in IQ means and, often, standard deviations.

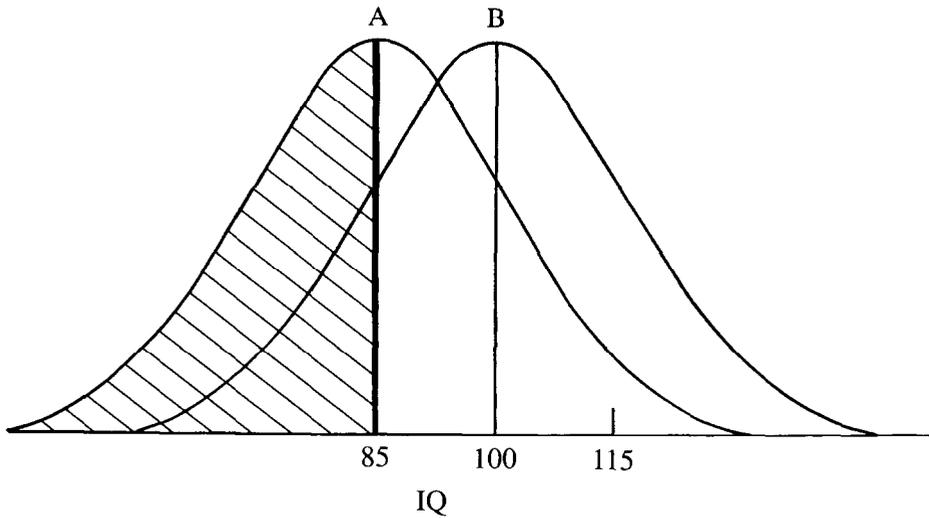
Students sometimes mistake critical IQs for the average IQs of persons experiencing the negative outcome, but this is clearly not the case. Neither is it the case that the parameters of the IQ distributions have necessarily been determined from the same individuals upon whom the prevalences are based. With suitable caution, group-level data from representative samples can be combined in the model with group-level data for a different variable from other samples representing the same population, because in all cases we are dealing with parameters, that is, statistics that distinguish one population from another. Synthesis of such data

enables analyses to be undertaken when data for IQ and outcome are not both present in the same sample, and is a distinct advantage, therefore, of population models. If inappropriate, such transfers of aggregate data are not likely to yield good results. If the assumption of normality of IQ distributions is too far off the mark, the model will not work well for yet another reason. The model itself, therefore, tests the tenability of the parameters applied and of the normality assumption as a by-product; other evidence concerning normality has been discussed elsewhere (Gordon, 1984, 1986).

As the Black and White prevalences in Table 1 and other tables to follow come in pairs for a given outcome, so do critical IQs. When the critical IQs of Blacks and Whites associated with often very different prevalences match or come close to matching (in view of the imperfections of real data), the difference in their prevalences is said to exhibit the property, *IQ commensurability* (Gordon, 1980d). This means that the difference between the two prevalences is closely commensurate with or consistent with the difference in Black and White IQ distributions, and hence the difference in outcomes can be understood as being entirely attributable to IQ. These aspects of the model are illustrated in Figure 1, which depicts perfect IQ commensurability in two hypothetical populations without going into the underlying specifics of the model. (Appendix A contains a more fundamental discussion of IQ commensurability and the mechanics of the model. It also distinguishes between two theoretical versions of the population–IQ–outcome model, called the proxy and the realistic models, that figure in the substantive interpretation of IQ commensurability.)

It is obvious, in Figure 1, that if the prevalences for Populations A and B were identical, say 50%, the distance between the two critical IQs would equal the distance between the means of A and B, in this case 15 IQ points. This would signal no relation between the large difference in mean IQ and the (zero) difference in prevalences. Intermediate degrees of fit are possible, but in this article only very close fits, differences in critical IQs of less than  $\pm 3$  IQ points, are interpreted as evidence of a relationship. The choice of what size difference to call a good fit is somewhat arbitrary, and a case could be made that even larger differences support the model. It should be evident that differences in critical IQs are simply goodness-of-fit measures for a between-group model that leaves open the exact shape of the probabilistic relation between IQ and the outcome within groups. In general, for negative outcomes like those examined here, the within-group probability functions will be gradually descending ones as IQ increases.

In national data, identical prevalences for Blacks and Whites would produce a difference in critical IQs of about 18 points, indicating total lack of fit because the prevalences failed to differ so as to reflect the 18-point difference between Blacks and Whites in mean IQ that reappears in major IQ surveys when all are converted to a common scale (Gordon, 1986, Table 2; see Table 2 here). Differences in regional data are usually somewhat smaller, about 16 points. Observed differ-



**Figure 1.** Example of perfect IQ commensurability: Critical IQs = 85 in both populations A and B 1.0 *SD* (15 points) apart in mean IQ, when Population A ( $M = 85$ ,  $SD = 15$ ) has a prevalence of 50% and Population B ( $M = 100$ ,  $SD = 15$ ) has a prevalence of 15.87%. The prevalences correspond to the percentages of the total areas under normal curves A and B, respectively, that lie to the left of the vertical line at A marking the critical IQs of 85. The interpretation of commensurability applies to any other location of matching critical IQs, such as IQ 100, marked by the vertical line at B, which would reflect still higher prevalences in populations A (84.13%) and B (50%). Even higher prevalences of 99.98% and 65.87% in populations A and B, respectively, would produce matching critical IQs at IQ 115.

ences in critical IQs can be interpreted with these indications of scale in mind and the knowledge that smaller differences indicate better fit.

It is crucial to make clear that when the population–IQ–outcome model fits prevalence data, the effects of IQ at the level of individual probands are not sufficient to produce IQ commensurability. A good fit demands, and hence reflects, the presence of contextual effects that are related to the IQs of their sources (such as parents and peers), effects that operate systematically so as to raise the probability or risk of a negative outcome for probands of a given IQ more in one population (the lower IQ one) than the other as a result of population differences in means and variances (as explained further in Appendix A). Consequently, a certain degree of vertical separation is required between the two within-group probability functions, reflecting the fact that the model incorporates IQ effects from each of the two prior levels, individual and contextual, and melds them together in a total IQ effect. It is this feature, above all, that provides a fundamentally new perspective on the role of intelligence in social matters. (Affinities between the population–IQ–outcome model’s interpretations in Appendix A and classic models of epidemiological contagion that may make the contextual aspects of the IQ model more understandable are described in Appendix B.)

### **Applying the Population–IQ–Outcome Model to Delinquency, Crime, Maternal Single-Parenthood, HIV Infection, and Poverty Prevalences**

Table 2 displays goodness-of-fit results from applying the IQ model to the populations and outcomes reported in Table 1. Small regional variations in IQ do occur, and these can be taken into account when possible, as in the case of Philadelphia in Table 2. The discussion begins with lawbreaking. Ranging from 0 to  $-2.9$ , numerous small Black–White differences in critical IQs indicate a quite satisfactory fit for delinquency and adult crime with one exception, on line 8. That exception, to be discussed in due course, illustrates a theoretical point.

*Delinquency.* Lines 1 through 7 in Table 2 pertain to juveniles, and the criteria of delinquency range over three grades of increasing seriousness: appearance in juvenile court, arrests for serious offenses only, and incarceration in a training school. The delinquency data on lines 1 to 4 (juvenile court record or incarceration) have been discussed extensively elsewhere (Gordon, 1976, 1986, 1987), and so only a few points need to be made here. The first is that fit of the model for the training school criterion is slightly poorer than for the court record criterion in Philadelphia, probably because the basic data were noisier. Blacks and Whites were not distributed equally over the 48 state jurisdictions involved in the training school data, and so variation reflecting differences in practice and policy from state to state was to be expected. Rural–urban differences in delinquency and racial distribution most likely also play a role here, but a smaller one than might be imagined, as the impression that urbanism is an important explanation of delinquency and crime depends mainly on confounding between size of place and racial composition (Gordon, 1976, pp. 253–256; Laub, 1983).

The results of substituting the means and *SDs* of socioeconomic status variables for IQ parameters in the model for the delinquency data on lines 1 through 4 have been reported elsewhere (Gordon, 1987).<sup>1</sup> If SES is truly as important as many commentators assume, then it would seem to be especially important in explaining juvenile crime. None, however, of numerous SES variables fit the data as well as IQ, and most performed very poorly indeed, yielding critical differences, when expressed in the same metric as IQ, of 5, 10, and even 20 or more. Differences in degree of fit among SES variables (e.g., income vs. years of schooling) were traced to differences in the degree to which the SES variables served as surrogates for IQ (see also Gordon, 1986). The best surrogate, the Duncan Socioeconomic Index (SEI) of prestige for all occupations held by the adult population (Duncan, 1961a, 1961b), was itself related by its author to the intelligence demands of occupations (Duncan, Featherman, & Duncan, 1972, p. 77). The SEI's smallest critical differences, scaled so as to be in the same metric as IQ, were 2.5 and 3.5 for the training school criterion in Table 2.

---

<sup>1</sup>This article (Gordon, 1987) was published with typographical errors in some tables uncorrected. Corrected copies are available from the author.

**TABLE 2**  
**IQ Commensurability of Black-White Differences in Prevalence of Juvenile Delinquency, Adult Crime, Single Motherhood, HIV Infection, and Poverty**

<b>IQ Means and Standard Deviations Used in Model</b>			
	<b>Blacks</b>	<b>Whites</b>	<b>B - W</b>
<b>Philadelphia parameters<sup>a</sup></b>			
<i>M</i>	86.0	101.8	-15.8
<i>SD</i>	12.5	16.4	
<b>United States parameters<sup>a</sup></b>			
<i>M</i>	83.8	101.8	-18.0
<i>SD</i>	13.1	16.4	
<b>Testing the Population-IQ-Outcome Model on Prevalences From Table 1: Implied Critical IQ and B-W Difference</b>			
	<b>Blacks</b>	<b>Whites</b>	<b>B - W</b>
<b>Philadelphia, prevalence of a juvenile court record by age 18.0</b>			
1. Boys	86.3	86.7	-0.4
2. Girls	73.5	71.8	1.7
<b>Nationwide, prevalence of incarceration in a training school by 18</b>			
3. Boys	60.9	63.7	-2.8
4. Girls	52.4	55.3	-2.9
<b>Philadelphia, arrests for UCR index offenses</b>			
Birth cohort I, 1945			
5. Males, by age 18	78.3	79.0	-.7
Birth cohort II, 1958			
6. Males, by age 17	78.0	79.7	-1.7
7. Females, by age 17	66.6	67.4	-.8
<b>Full lifetime arrest prevalence, nationwide, males</b>			
8. All nontraffic	90.0	100.5	-10.5
9. UCR index offense	84.5	84.5	0.0
<b>Incarceration prevalence, nationwide (approximated from NLSY); incarcerated when interviewed</b>			
10. Males, mean age 29	69.1	69.4	-0.3
<b>Males under criminal justice system supervision</b>			
11. 1989, age 20-29	74.1	76.6	-2.5
12. 1994, age 20-29	77.0	77.2	-0.2
<b>Algebraic mean of 1-7, 9-12<sup>b</sup></b>			<b>-.96</b>
<b>Prevalence, nationwide, of one-parent, mother-headed families</b>			
13. 1970	78.0	79.7	-1.7
14. 1980	83.4	84.9	-1.5
15. 1990	85.8	87.3	-1.5
16. 1993	86.6	88.1	-1.5
<b>HIV infected as of January 1, 1993</b>			
17. Males, aged 27-39	59.2	61.7	-2.5
18. Females, aged 27-39	53.4	48.7	4.7
19. Both sexes, aged 27-39	57.0	58.2	-1.2
20. Males, aged 18-59	57.6	59.4	-1.8

(continued)

**TABLE 2**  
**Continued**

	Blacks	Whites	B - W
21. Females, aged 18-59	51.9	47.8	4.1
22. Both sexes, aged 18-59	55.4	56.2	-.8
<b>Children living in poverty, nationwide</b>			
Official poverty			
23. 1939	107.9	110.1	-2.2
24. 1949	99.0	98.6	.4
25. 1959	88.9	88.1	.8
26. 1969	82.1	82.2	-.1
27. 1979	79.7	82.9	-3.2
28. 1979 (CPS)	80.8	82.4	-1.6
29. 1988 (CPS)	81.8	85.0	-3.2
Relative poverty			
30. 1939	92.3	95.3	-3.0 <sup>c</sup>
31. 1949	89.6	88.9	.7
32. 1959	88.2	87.3	.9
33. 1969	85.7	86.5	-.8
34. 1979	83.7	87.5	-3.8
35. 1979 (CPS)	84.8	86.7	-1.9
36. 1988 (CPS)	84.7	89.3	-4.6
<b>Families with children living in poverty, of all families with children</b>			
Official poverty			
37. 1939	105.9	107.0	-1.1
38. 1949	96.4	95.8	.6
39. 1959	86.1	86.1	0.0
40. 1969	80.0	80.7	-.7
41. 1979	78.5	81.8	-3.3
42. 1979 (CPS)	79.3	81.2	-1.9
43. 1988 (CPS)	80.5	83.7	-3.2
Relative poverty			
44. 1939	90.3	92.3	-2.0
45. 1949	87.4	86.7	.7
46. 1959	85.5	85.2	.3
47. 1969	83.5	84.6	-1.1
48. 1979	82.3	86.2	-3.9
49. 1979 (CPS)	83.3	85.3	-2.0
50. 1988 (CPS)	83.3	88.0	-4.7

*Note.*<sup>a</sup>IQ parameters are ones employed before, based on an earlier review of major IQ studies (Gordon, 1986, Table 2, 1987, Table 1), and prior work (Gordon, 1976). For the U.S. as a whole, the reviewed studies reveal an average Black-White difference, when all test data are converted to the common metric of the 1937 Stanford-Binet normative sample (Terman & Merrill, 1960), of 1.1 White *SD*, or 18 IQ points. A later review of IQ studies (Herrnstein & Murray, 1994, p. 277) also found a 1.1 White *SD* Black-White difference. Regional IQ parameters are used in analyses of all Philadelphia prevalence data. The national Black/White ratio of *SD*, .8, is based on convergence in three major studies with very large samples. Later national standardization studies, with much smaller Black samples of about 200-300, have yielded Black/White *SD* ratios of .906, .888, and .987, the last for a relatively unfamiliar test (respectively, Kaufman & Doppelt, 1976, Table 4; Reynolds, Chastain, Kaufman, & McLean, 1987, Table 1; Lynn, 1996, Table 1). Addition of these discrepant values to a weighted average Black *SD* worsens the fit, usually, only of prevalence rates for Blacks that are 4% or lower (e.g., lines 3, 4, and 17), but sometimes improves it (lines 18 and 21). Lines 19, 20, and 22 continue to fit either way.

<sup>b</sup>Line 8 was included only to illustrate a problem with definitions of offending that are too inclusive of mild offenses.

<sup>c</sup>In view of its location on the borderline defining good fit, let it be noted that this difference, with rounding error corrected, was -2.9.

Newly analyzed prevalences of delinquency on lines 5 to 7 in Table 1 pertain to a different criterion (police contacts, presumably arrests, for serious offenses) and were obtained by a different method. Lines 1 through 4 were obtained by the synthetic cohort method using cross-sectional age-specific rates of first qualifying as delinquent to calculate lifetime prevalence (Gordon, 1973; Gordon & Gleser, 1974); lines 5 through 7 were determined from longitudinal data by following birth Cohorts I and II retrospectively, each born in a different year. Both methods happen to be applied to the same jurisdiction, Philadelphia (lines 1 and 2 and 5–7 of Table 1). And both methods yield equally good fits for their different criteria, as Table 2 shows. (Implications of the equivalence of the two methods are discussed in Appendix C.)

**Adult Criminality.** The several entries following the delinquency data in Table 2 successfully extend the model to criminality by adults, and to points in time many years later. The model succeeds by showing IQ commensurability for serious offenders (entries 9–12), but a single entry, on line 8, has been inserted to illustrate an auxiliary principle that covers the disruptive effect on fit of including deviance that is too mild. This is the deviance principle, which recognizes that the degree of mistake represented by a deviant outcome can moderate the applicability of the population–IQ–outcome model. (Observations leading to formulation of the deviance principle are described in Appendix C and at a later point.)

On line 8, full lifetime prevalence based on all nontraffic (i.e., many mild) arrests fails to fit by a wide margin; too many Whites meet this inclusive criterion, or too few Blacks, given their two IQ distributions. Conceivably, mild offenses are reported relatively more often in White neighborhoods, but ignored or tolerated in Black neighborhoods, thus accounting for the disparity. Police contacts, for example, often arise from telephoned or mailed complaints from civilians, and most such contacts do not result in arrest (Wolfgang, Figlio, & Sellin, 1972, pp. 43, 221). The overrepresentation of Whites implied by line 8 is typical of outcome criteria that incorporate mistakes that are relatively minor.

On line 9, a test of prevalence data from the same study as the data on line 8 reveals, now instead of the worst fit in Table 2, a fit that is perfect. The difference is consistent with the difference in the seriousness of the two criteria. Line 9 is based on Uniform Crime Report index offenses, which the FBI uses to measure serious crime.

Line 10 reveals a near-perfect fit to the IQ model for data from *The Bell Curve*, based on incarceration, a criterion that reflects serious criminality. Fischer et al. (1996, pp. 93–94, 254) have criticized Herrnstein and Murray's (1994, p. 249) within-group finding that intelligence was far more important than parental SES background in accounting for incarceration. Fischer et al. base their argument on regression models incorporating variables that not only are different from parental background status but that serve as predictable correlates of probands' own intelligence. Examples of such variables are academic school track, quality of students

at the school attended, and years of schooling completed when intelligence was tested at an average age of about 18 or 19. Such variables, which pertain to the proband, are not normally counted as part of parental SES background; in addition, they substantially dilute the effect of the Armed Forces Qualifying Test (AFQT) used to measure  $g$ , just as would inclusion of other IQ tests under different names. Unique names for variables often obscure common properties.

The between-group test of the model on Herrnstein and Murray's prevalence data in Table 2 does not employ the AFQT scores, but instead relies, like the other tests of the model, on IQ parameters determined from other studies, which would reflect the IQ contexts of the probands better than these AFQT scores. The fit of *The Bell Curve* criminality data on line 10 is consistent with many other findings in Table 2 concerning the role of  $g$  in between-group differences in criminality. Unlike regression models fitted by least squares, a good fit for the population–IQ–outcome model requires a specific amount of criminality in one race, given the amount in the other race and the model's IQ parameters. The precise fit attained from intelligence alone, which left a critical difference of only  $-0.3$  IQ points, suggests that the dismissal of intelligence by Fischer et al. was overhasty. Despite a  $.86$  *SD* race difference in observed parental SES (Fischer et al., 1996, Table A2.1), no other variables besides IQ are required to account for the difference in prevalence, as IQ does so quite successfully, more successfully, in fact, than SES. If observed parental SES is substituted for IQ in the model (following Gordon, 1987), the critical Black–White difference, expressed in the same metric as IQ, turns out to be  $-3.2$  instead of, as in Table 2,  $-0.3$  for IQ. That the fit of SES in this case is not worse is owed to the fact that it is a composite variable including several good surrogates for IQ in the parental generation: mother's and father's education and the Duncan SEI for the higher of their occupations.

Entries 11 and 12 demonstrate another two good or nearly perfect fits for adults whose offenses were severe enough to warrant major involvement with the justice system. The Sentencing Project (Mauer & Huling, 1995), the source for the 1989 and 1994 entries, attributed the greater number of Blacks under supervision of the criminal justice system mainly to racial disparities in class and economic well-being, and the greater increase for Blacks than Whites between 1989 and 1994 (Table 1, above) to disparate impact of the War on Drugs (see also DiMascio, 1995, p. 28; Tonry, 1995, chap. 3).

Table 2 reveals, however, that the race differences in question were closely commensurate with IQ differences all along and that the greater change from 1989 to 1994 for Blacks, which appears to be mainly the effect of changes in sentencing policy for drug convictions rather than of new convictions (e.g., Langan, 1991, Table 3), simply brought their rates into closer alignment with the IQ model. The observation concerning improved fit to the IQ model does not affront the explanations offered by others so much as supplement them by pointing to a deeper cause.

Differences between critical IQs can be averaged to summarize the success of

the model in multiple applications within the same domain. Not counting line 8, because it was included only to illustrate the inappropriateness of prevalences based on criteria that incorporate too many mild offenses in their definitions, the mean difference for 11 remaining applications to the prevalence of crime and delinquency is only  $-.96$  IQ points (Table 2). For over 45 years, therefore, the difference between Blacks and Whites in varying levels of prevalence of the more serious forms of crime and delinquency has remained closely consistent with the difference in their IQ distributions.

***Single Motherhood.*** The percentage of one-parent families has long been regarded as a factor contributing to rates of juvenile delinquency (Lykken, 1995; Miller, 1958; Moynihan, 1965; Wilson & Herrnstein, 1985, chap. 9), but the link, if any, has not gone uncontested (e.g., Berger & Simon, 1974; Blumstein, Cohen, Roth, & Visher, 1986, pp. 45–46; Hirschi, 1969, pp. 242–243; McCord & McCord, 1959; Robins, 1966; Rosen & Neilson, 1978). If, however, both delinquency and single parenthood have a common cause, the observed association between the two variables might be partly spurious (noncausal), yet partly causal too. Single parenthood could often coincide with, rather than cause, delinquency, but simultaneously it could also serve as a vehicle or marker for contextual effects of  $g$  that contribute to delinquency (e.g., poor parenting by those who become teenage mothers) but are relatively weak compared with effects of the delinquent's own  $g$ . Such chaotic sources of association could account for the frequently weak or conflicting results obtained when testing whether family breakup and father absence contribute to delinquency (e.g., Wilson & Herrnstein, 1985, p. 247). Neighborhoods and housing projects with high delinquency rates do tend to have high rates of out-of-wedlock births, a major cause of single motherhood, but this could signal only that both outcomes result from a common cause (Gallagher, 1996, pp. 48–49). Evidence indicates that cause is  $g$ .

Lines 13 through 16 in Table 2 reveal that Black–White differences in rates of single-parent families, with children under 18 and maintained by mothers in 1970, 1980, 1990, and 1993, are commensurate with Black–White differences in IQ distributions, just as were rates of delinquency and criminality. The recurrent negative difference of about  $-1.5$  IQ points between critical IQs of Blacks and Whites suggests that prevalence of such families may actually be somewhat higher among Whites than among Blacks once IQ distributions are taken into account. Herrnstein and Murray (1994, p. 175) found divorce to be inversely related to IQ among Whites, but also that higher parental SES enabled young White couples to become divorced more easily when their IQ was held constant. This added SES effect may help explain the relative excess of Whites, among whom divorce accounts for a much larger share of children in mother-only families than among Blacks (Hernandez, 1993, Table 3.6).

Prevalence of single-parent families maintained by mothers did increase over

the decades in question for both races, more so for Whites than Blacks by one method of calculation (a factor of 2.3 vs. 1.8, 1993/1970), as is often remarked, but the increases maintained their relation to IQ, as though the outcome was becoming less deviant and hence less of a mistake in relative terms (although still representing a major life commitment). Recall, the model accounts for relative, not absolute, rates, especially when the outcomes have not arrived at equilibrium, as single motherhood evidently has not.

Intelligence could figure in single parenthood through its effects on management of sexuality, contraception, mate selection, marital adjustment, and availability of marriage-minded individuals (e.g., Roth, 1994, chap. 10). Recall that rates of out-of-wedlock births, an important component of rates of single parenthood, decline as women's IQ increases (Berlin & Sum, 1988, Figure 9; Herrnstein & Murray, 1994, p. 183). Lykken (1995, pp. 219–220) has suggested that the sizable proportion of young adult Black males who are incarcerated at any time helps drive Black illegitimacy rates by lowering the male:female sex ratio and thus undercutting motivation to bond exclusively with one woman. But we have already seen that race differences in incarceration rates are themselves commensurate with IQ, so even if Lykken's hypothesis is correct, the explanatory status of IQ has not necessarily been compromised. Instead, we have a potential example of one of the many paths by which population IQ can affect various outcomes independently of the IQs of probands. Another such path may be through the poor parenting provided by many unmarried single mothers. Lykken (1995, pp. 214–216) uses Black–White differences in rates of unmarried single motherhood to account for Black–White differences in rates of chronic offenders, but both sets of rates may be related to IQ, as the data presented in Table 2 indicate.

Judging from the results in Table 2, achieving and maintaining a socially desired family structure appears to function like a weak IQ test within the Black and White races despite changes in the absolute level of the outcome over the course of three decades. Differences in family structure between the Black and White races, on the other hand, are closely consistent with population IQ differences mediated through a variety of variables ranging from employability to imprisonment.

Analyses of related data (not presented here) provide additional historical perspective on these results, linking them, in particular, to secular changes in attitudes toward out-of-wedlock births. Race-specific census data for female-headed families as a percentage of all families are a related but somewhat different family structure outcome that is less apt for present purposes (female-headed families need not include children). However, such data extend back to 1940 and are conveniently available by individual years from 1970 to 1983 (Wilson & Neckerman, 1986, Table 10.1). Black–White differences in critical IQs absolutely smaller than  $-3.0$  appear in this series from 1972 on, only a bit later than rates for the

single-parent families maintained by mothers beginning at 1970 in Table 2. The differences in critical IQs tend to diminish in size over time, but are always negative. This series, therefore, falls in line with the data in Table 2, both in respect to approximate timing of the onset of fit and the negative difference.

Race-specific rates of children aged 0 through 17 in mother-only families (Hernandez, 1993, Table 3.6) (note, based on counting children, not families), another related statistic (not presented), show poor fits in 1940, 1950, and 1960, but acceptable and still improving fits in 1970, 1980, and 1988.

The timing of the point at which the model begins to fit in these various series, circa 1970, coincides with a weakening of attitudes against, and of legal barriers to, out-of-wedlock pregnancies (Murray, 1984; Roth, 1994, chap. 10). In the late 1960s, for example, federal legislation prohibited preferences in favor of married couples in low-income public housing, and in 1972 it became illegal for school systems receiving federal monies to expel students or to restrict their activities because of pregnancy (Gallagher, 1996, pp. 253, 255). In one study, pregnancy was the reason most cited by female teenagers for dropping out of school (Alan Guttmacher Institute, 1976, p. 25). Use of the term *illegitimacy* fell into disfavor. These policy changes, probably adopted in part because the preexisting ones had an adverse impact on Blacks and on the schooling of Blacks during a time of special attention to poverty and race, tended to normalize unwed motherhood. Marriage in general was no longer as well supported by institutional means (Akerlof & Yellen, 1996); this left matters more subject to unassisted individual discretion.

***HIV Infection.*** Epidemiologists know of no good explanation for race differences in prevalence of HIV infection in the United States, although low education and poverty are often mentioned. A greater concentration of infected persons in one race than another initially and the proportionality of the velocity of spread to that initial concentration can be invoked to explain the greater prevalence today, but that leaves unexplained the difference in initial conditions. The data in Table 2, as will be explained, can provide an explanation of both initial and present differences in prevalence based on population differences in intelligence.

Table 2 displays fit for six prevalences of HIV infection (entries 17–22). The sex-specific data fit the IQ model more poorly for females than males (critical IQ differences of about 4 vs.  $-2$ ). If the two sexes are combined, the overall fit becomes better than for either one alone, and in fact quite good (about  $-1$  on lines 19 and 22). Recall from Table 1 that the combined-sex infection prevalence for Blacks was about five times that for Whites. This racial differential in HIV infection is greater than the differentials for crime prevalences, and yet the population differences in IQ are commensurate with the racial disparity in prevalences in both cases. It is impressive that knowledge of the prevalence for Whites on line 22 of Table 1, for example, together with the IQ parameters of both races would enable someone using the model to predict the prevalence for Blacks to within less than

one quarter of one percentage point. Starting instead with the prevalence for Blacks, one could predict the prevalence for Whites within .05 percentage points.

In the spread of HIV, each sex is very much a key part of the social context of the other, whether for sex or needle sharing, perhaps far more so than in the case of crime and delinquency, where "the associates of young persons in delinquency are almost always of the same sex" (Reiss, 1986, p. 132). The combined rate for HIV may do better justice to this fact than the sex-specific rates. Because of the large sex difference in prevalences, males in each race may constitute a more relevant HIV context for females than females at present do for males. About 70% to 80% of persons infected through heterosexual contact are women (Holmberg, 1996, p. 645). It is of interest, too, that data for the broader age range in Table 2, which includes the data for the narrower age range, provide the better fit. Perhaps the larger universe afforded a more reliable sample size, as well as a longer opportunity for the effect of IQ to play out.

Choosing a sex partner or needle sharer means not only going to bed in effect with all of that person's previous sex and needle-sharing partners, as AIDS education teaches; it means accepting the judgments concerning lifestyle, partners, and precautionary practices as well as the sexual choices available throughout that person's network of partners, many remote from the proband. The collective intelligence of the network thus figures in the outcome of every individual decision. Collective behavior is not miraculously decoupled from collective intelligence. As spokeswoman Gina Bianco for the Delaware Division of Public Health put it, "For the most part AIDS is a disease contracted by risk-taking behavior" (Spizzerrri, 1996, p. A2). There is no low-IQ-related differential detection hypothesis for HIV as there is for crime (as discussed in Appendix C), only a differential contraction hypothesis.

Access to and utilization of information are thought to play a major role in controlling HIV (Philipson & Posner, 1993). The importance of information helps to explain why differences in  $g$  between two sexually segregated populations might produce systematic differences in outcomes that seem totally explainable by the populations' distributions of  $g$ , even though  $g$  may be far from the total explanation of the same outcomes at the individual level within each population, as we know (consider, for example, hemophiliacs; Meier, 1996). Group differences between populations in how quickly they assimilated knowledge about HIV, that is, how easily they became educated about the basic facts of the disease, once that information became public, could conceivably account for an initial difference in prevalence that partly or totally accounts for the present difference; there is every reason to expect, however, that the effects of  $g$  on the epidemic are ongoing rather than one-time events.

The spread of HIV infection is quintessentially a problem that invites social network analysis (e.g., Dyson, 1991; Kretzschmar et al., 1994; Orubuloye, Caldwell, Caldwell, & Santow, 1994), but  $g$  is not normally viewed as a vital strand in

such networks. Failing to consider the IQ commensurability of HIV prevalence and measuring instead education (where Black–White differences in median years of schooling have all but vanished in recent U.S. cohorts) could complicate the search for good epidemiological models of the spread of HIV (e.g., Kaplan & Brandeau, 1994) and obscure a relevant risk factor.

***Poverty Rates (Prevalences) for Children and for Their Families, Absolute and Relative.*** Because poverty is cited frequently as a major cause of Black–White differences in other important outcomes, ranging from deaths from breast cancer and tuberculosis to learning difficulties and crime (e.g., Fischer et al., 1996, pp. 194–196; Huber & Chalfant, 1974; Hurley, 1969; Tonry, 1995, p. 9; Watson, 1992, p. 57), it is of special interest to explore that variable with a model that may also help to resolve a fundamental question about the nature of poverty as an explanation.

Is poverty to be understood as a continuous variable that is measurable, or as a virtually unanalyzable qualitative state so global that no set of measured variables seems to capture it adequately? According to the first view, “poverty is most simply and clearly understood as a lack of money” (Ryan, 1971, p. 117), a conception of the variable known as “*income poverty*” (Plotnick & Skidmore, 1975, p. 6). Improvements in specified variables such as income, however, often leave the dysfunctional behaviors they supposedly help explain, the so-called culture of the poor, little changed. Disappointments with variables that seem potent otherwise, such as income, years of schooling, and job training, have given rise to a more pessimistic, qualitative view of poverty.

The qualitative conception of poverty, “the seemingly intractable urban slum” (Plotnick & Skidmore, 1975, p. 4), is documented in statements such as the following: “action at any one point on the poverty cycle would be useless without action at every point, to break the hold of an entire way of life”; poverty program participants could not be aided successfully “until the whole culture [to which they returned] was transformed” (Zarefsky, 1986, p. 108); “no spectacular breakthrough can be made until the whole structure of the culture of poverty is destroyed” (Hurley, 1969, p. 73). Such an indivisible qualitative state seems to demand an equally qualitative explanation, yet none has been suggested other than poverty itself. This recourse to poverty as its own explanation constitutes the cycle-of-poverty theory; born of desperation, the theory is perhaps no more than a thinly disguised tautology, and it fails to account well for why some individuals and groups have emerged from a history of impoverishment and others have not.

The effects of general intelligence and the contexts it gives rise to may have the requisite pervasiveness to account for the widespread impressions of discontinuity and qualitiveness concerning poverty, while at the same time restoring continuousness and measurability to the explanatory variable.

Of the various poverty rates available and tentatively considered, I have settled

on rates for children for several reasons. First, that choice imposes some degree of age standardization across race by excluding family units composed solely of elderly persons, whose poverty status may be related, moreover, to the special health and employment problems associated with aging as well as to certain welfare advantages enjoyed by the elderly and other complications (Ruggles, 1990, pp. 71–72, 169). Second, data derived from a reasonably uniform treatment of both absolute and relative poverty rates, spanning a series of years (including historical years not previously includable) are available for children by race from a major study by Hernandez (1993, Tables 9.4 and A-9.4). His inclusion of the early dates 1939 and 1949, as it happens, will prove to be especially valuable for documenting a trend. Third, unlike some poverty rates (e.g., for children in mother-only families), race-specific rates for all children should reflect populations reasonably representative of entire IQ distributions, as entire distributions change little in IQ between adjacent generations. Fourth, poverty in childhood is often thought, rightly or wrongly in specific situations, to have a more formative influence on later development than poverty in adulthood; indeed, within the “cycle of poverty” (Levitan, 1969, p. 281) and “culture of poverty” (Lewis, 1966) perspectives, unameliorated childhood economic disadvantage is held to lead to adult economic disadvantage.

Poverty researchers distinguish and employ two major kinds of poverty rates: *absolute* and *relative* (Hernandez, 1993; Ruggles, 1990, chap. 2). Tables 1 and 2 (entries 23–36) report and test seven prevalences for each. Absolute poverty rates (known also as official rates because used as the official U.S. standard) employ a threshold that remains fixed over time when expressed in constant dollars. Although arbitrary to a point, such a threshold reflects the judgment of experts concerning the needs of a traditional family of four, and is adjusted to apply to other family configurations while taking account of economies of scale in larger families (Hernandez, 1993, p. 242; Orshansky, 1969; Ruggles, 1990, pp. 64–67).

Absolute thresholds contain the possibility that poverty so defined could decrease markedly over a period of decades of rising real income, perhaps to vanish altogether. Thus, “the three decades following the Great Depression brought large increases in absolute income and large reductions in absolute want,” and the official poverty rate for children dropped from 72% in 1939 to 16% in 1969 and 1979 (Hernandez, 1993, p. 238; see also Herrnstein & Murray, 1994, chap. 5; and Table 1 here). An unspoken consideration, perhaps, is that if poverty so defined declines drastically, but the phenomena it is purported to cause remain stable or increase, the explanatory status of poverty is thrown into question. Herrnstein and Murray (1994, p. 142) raised this implication when they pointed out that “poverty itself has been declining as various discontents have been rising during this century.”

The concept of relative poverty can help circumvent the problem raised by declines in absolute poverty that fail to be reflected in concomitant declines in many of its putative consequences. Decreases from 1939 to 1988, for example,

are less drastic for relative than for absolute poverty. Relative poverty shows fewer persons in poverty in 1939, and more persons in poverty in 1988, than absolute poverty does. As operationalized in the rates employed here, the relative poverty threshold was set at the commonly recommended level of 50% of the median income at any given time (Hernandez, 1993, p. 235).

One of the noteworthy features of relative rates is that they cannot decline unless changes occur in the shape of the income distribution as a whole (Ruggles, 1990, p. 19). A peculiarity of relative poverty rates is that they may well decrease in times when real income is falling and increase when real income is rising. Many key variables that might conceivably be implicated in absolute poverty, such as malnutrition, cannot be assumed to be equally implicated under the rubric of relative poverty; after a point, how well one eats has little relation to physical requirements. Unfortunately, use of the single term *poverty*, with all of its poignancy, for both kinds of rate is not conducive to maintaining such distinctions (relative deprivation would seem a better term for relative poverty, although it would not have the same emotional impact). Many authors favor relative over absolute definitions of poverty because the state of poverty has, they point out, a strong normative aspect at any given time in any cultural context (Hernandez, 1993, pp. 238–244; Ruggles, 1990).

Lines 23 to 26 in Table 2 reveal that at four widely separated points in time, beginning in 1939 and ending in 1969, the population–IQ–outcome model accounts for Black–White differences in official (absolute) poverty prevalences of children very well, often almost perfectly. Three of the differences in critical IQs are less than 1 in absolute value, and the algebraic average of the four differences is only  $-.3$ , a remarkable outcome. Impressive too is the fact that the model successfully tracks the poverty prevalences through major changes in level during the 1939 to 1969 period; rates for Blacks and Whites decreased by 51.9 and 57.8 percentage points, respectively, and the Black/White ratio of rates increased dramatically from 1.4 in 1939 to 3.9 in 1969. Reliance on such Black/White ratios would lead to the conclusion that the relative position of Blacks worsened, although the model indicates that Blacks and Whites maintained IQ commensurability throughout a period of drastically declining poverty prevalences.

Starting in 1979 (lines 27–29 in Table 2), the fit deteriorates markedly, even allowing for some inconsistency in the two estimates from different sources for 1979. From another source, detailed data by individual years from 1974 through 1992 (not presented) indicate that the model successfully fits Black–White differences in official poverty rates for children under 18 years of age from 1974 through 1981 (i.e., 2 years beyond 1979), although, by showing acceptable but large critical IQ differences of  $-2.7$ , the years 1980 and 1981 foreshadow the generally poor fit from 1982 to 1992 (there was briefly some improvement during 1987–1989) (U.S. Bureau of the Census, 1993, Table 3).

Relative poverty prevalences for children on lines 30 through 36 behave in a

manner closely parallel to that of absolute rates on lines 23 through 29, even with respect to sign. For the good-fitting years 1939 to 1969, the algebraic average of the four differences in critical IQ is only  $-.5$ , just slightly larger than that for absolute poverty.

Children are not placed into poverty through IQ-dependent actions of their own, obviously, and so cannot be the relevant actors, yet children are the units of analysis for the poverty rates tested thus far. When the model fits those prevalences, it seems to anticipate that rates for children in poverty will translate into IQ-commensurate rates of households or parents in poverty, where the intelligence of adults rather than of children could account for the outcome. Such an effect of IQ, interestingly enough, would be entirely a contextual one on children, and so it dramatizes how important the intelligence context can be. Variation in family size prevents prevalences for children from serving as direct substitutes for prevalences for families, however. Poverty rates for families proper are needed that apply to the very same families on which the rates for children were based.

Fortunately, race-specific data on “dependent sibsize distributions” (percentages of children 0–17 in sibships of various sizes) and poverty rates by sibsize make it possible to convert the poverty rates for children into poverty rates for the families containing those children (Hernandez, 1993, Tables 9.4 and A-9.4). (Hernandez’s Table 2.4 was used to calculate mean sibsizes within sibsize categories of his Tables 9.4 and A-9.4. Dependent sibsize distributions, which are for individuals, not sibships, were divided by mean sibsizes to obtain sibsize distributions for families. These distributions were weighted by poverty rates by sibsize and the products summed to derive the final poverty rates for families, keeping in mind that each sibship represents a family.) Results of testing the model on the derived rates for families are displayed on lines 37 through 43 and 44 through 50 for both absolute and relative prevalences, respectively.

The poverty picture is essentially unchanged by the conversion to prevalences for the families. Again, the model fits from 1939 to 1969, yielding algebraic averages of critical IQ differences for absolute and relative poverty equivalent to those based on prevalences for children. From 1979 on there again appears a consistent overrepresentation of Whites (or underrepresentation of Blacks) in poverty according to the model. During the decade of the 1970s something had evidently intervened to disturb the fit of the model by the early 1980s, some factor that affected one race more than the other. Mutually supportive historical and statistical data suggest what that factor was, and suggest further insights into certain failures of the model.

Zarefsky’s (1986) history of the Office of Economic Opportunity (OEO) revealed that by the late 1960s the War on Poverty was becoming viewed, even by the White poor, who consequently came to reject much of what OEO had to offer, as a program designed for Blacks and aimed principally at the ghetto (pp. 104–105). “Most OEO programs were focused on the ghetto” (p. 103), in part, because

that was where “the greatest *concentration* of poor people was to be found” (p. 103), and perhaps in part because OEO efforts were viewed often as a means of preventing then-recurrent urban riots (pp. 111–119, 170). “By 1967 . . . the rate of increase in the population receiving Aid to Families with Dependent Children had doubled,” because “early community action programs . . . organized the poor with the goal of securing rights already recognized. As a consequence, these organizations began to advertise the availability of public welfare” (p. 150). Community action programs were OEO’s largest single budget item (Plotnick & Skidmore, 1975, pp. 22–25, Table 1.2). In 1985, eligible Whites were still viewing poverty programs as being intended mainly for Blacks (Orfield & Ashkinaze, 1991, pp. 185–187).

Beginning in 1969, statistics reveal that the contribution of government welfare to the reduction of official poverty among children from the level determined by father’s income alone became appreciably larger among Blacks than among Whites (Hernandez, 1993, Table A-9.3). Gottfredson’s (1986a, Table 2) data revealed that substantial employment (and thus, by implication, income) gains unrelated to IQ qualifications were made by Blacks as compared with Whites during the 1970 to 1980 decade, presumably under the mantle of affirmative action following the *Griggs v. Duke Power Co.* decision in 1971, which codified disparate impact of employment standards as the definition of racial discrimination (Sharf, 1988). Gottfredson’s results were later confirmed by Herrnstein and Murray (1994, pp. 321–322): “The chance of entering a high-IQ occupation for a black with an IQ of 117 . . . was over twice [that] of whites with the same IQ.”

These racially differential historical and statistical effects are all in a direction to skew the fit of the model in the manner already described by the time that it ceases to fit prevalences of poverty well in 1979 or thereabouts. Recall that it was during the same tumultuous decade that single-motherhood prevalences commenced, rather than ceased, to fit the model well. Parenthood underwent a diminution of social intervention at about the same time that poverty was undergoing an augmentation in societal intervention, and consequently the two trends in fit run in opposite directions.

The fact that for some variables successive observations during one period fit the model better than during another period, as in the cases of single motherhood and poverty, when accompanied by a plausible explanation for the change, does not mean that IQ ceases to operate once the fit worsens. Instead, periods of good fit should be regarded as affording invaluable glimpses into the possible role of intelligence on group differences in outcomes that may often be obscured by other, perhaps relatively minor, influences on prevalence differences, such as those described for single motherhood and poverty. Outcomes exhibiting such systematic changes in fit testify to the basic sensitivity of the model to small disturbances when goodness of fit is fairly stringently defined. By contrast, the absence of comparable disturbance to outcomes that persist in fitting over a wide span of years, such as

criminality, provides theoretical, if not sobering, insight concerning their robustness in the face of traditional interventions aimed at reducing racial disproportions. Disturbances to the fit for poverty, for example, did not result in disturbances to the fit for lawbreaking during the period in question (see Tables 1 and 2). Although it has been proposed that the most commonly used theory to explain the race-crime link is poverty ("Summary," 1990), as late as 1994, the model was still fitting criminality prevalence data well.

### THE LEVEL OF ENTIRE POPULATIONS—OPINION OUTCOMES

This major section extends the IQ model into the domain of public opinion concerning what is to be accepted as factual. The domain of factual opinion often lacks the constraints imposed on performance outcomes by real consequences, and it adds the element of irreducible uncertainty that characterizes events not personally witnessed by all. Such considerations can work for or against success of the model. Uncertainty calls forth judgment, but loosening of practical constraints can add unpredictability to outcomes.

Surveys concerning belief in conspiracy rumors and key beliefs about the O.J. Simpson trial provide the main data.

#### **Conspiracy Rumors in Everyday Life**

Although conspiracy themes among Whites have long been studied (Cohn, 1966; Harrington, 1996; Hofstadter, 1965), the many rumors centered around conspiracies, some quite elaborate, that have been afloat in the African-American community at one time or another have just recently drawn scholarly attention (Turner, 1993). Often, those rumors concern consumer products manufactured by White corporations and their imagined sponsorship by the Ku Klux Klan (e.g., *Kool* cigarettes because of the K) or supposed adulteration of fast foods aimed at sterilizing Black males. Others target the government. Black celebrities, authority figures, academics, and media outlets sometimes lend credence to the rumors, and peers sometimes punish better informed individuals for defying the messages (DeParle, 1990; Fiske, 1996, chap. 4; Fletcher, 1996; Turner, 1993, pp. 132–135, 188). In Washington, DC, such rumors are said to be widely known among Blacks, no doubt often with wry humor, as "The Plan" (Cooper, 1996, p. A4; DeParle, 1990, p. B6). The same term has been reported in use for "virulent conspiracy theories" among Puerto Ricans and Blacks living in Manhattan (Bourgois, 1995, p. 50).

Some students of these rumors see them as functional to the Black community, perhaps in practical ways, such as in steering consumers away from pricey athletic shoes, but especially in sociological ways, by contributing "to an atmosphere of communal problem-solving," serving as "tools of resistance," and creating a

common culture from within which leadership, mass mobilization, and concerted action can arise (Turner, 1993, pp. xvi, 127, 164; see also Fiske, 1996, chap. 4).

Contrary to the benign view of the functions of such rumors taken by folklorists and some Black activists, others see rumors (like those in Table 3) as playing a seriously dysfunctional role even in the Black community. Black organ donations are few because of distrust of the medical system and belief in myths ("Why More," 1995), fostered, for example, by Nation of Islam leader Louis Farrakhan, who stated that Black-on-Black homicide is favored by Whites because it makes organs available for transplant to Whites (Crawford, 1994). Black physicians and others denounced Farrakhan's claim as harmful to Blacks, and disregarding of the facts, pointing out, "It is mostly Caucasian organs that are transplanted into African American patients, not vice versa" (Crawford, 1994, p. A15). Blacks, not Whites, are overrepresented on waiting lists for kidneys, for example, by a factor of three.

"Public-health workers say that the discussion of AIDS as a plot against blacks has eroded the credibility of AIDS prevention campaigns and has made some blacks suspicious of AIDS tests and treatments" (DeParle, 1990, p. B6). One AIDS worker found that it made her job harder, "Because I have to devote time and energy to addressing those concerns. If I dismiss it out of hand, then those people who believe in it have turned me off" (DeParle, 1990, p. B6). That both higher HIV prevalence and more frequent endorsement of beliefs that obstruct control of the AIDS pandemic should be found within the same population is sadly ironic, but consistent with the pervasive and often multiple co-endogenous (parallel) effects of *g*.

Others spoke more generally about effects of such rumors. The director of a Black studies program stated that they diverted attention "from the resolution of pressing social problems" (Myers, 1990, p. A19). Andrew Cooper, publisher of a Black newspaper in Brooklyn, warned that the rumors were dysfunctional for all, observing, "It is a danger to America to have a large group of its citizens believe that its government is in a conspiracy to eliminate them" (DeParle, 1990, p. B6).

A Black debunker of conspiracy rumors submitted that their impact "is simply more potent than any empirical evidence a researcher can supply" (Turner, 1993, p. 214). Rather than being uniquely functional to any one community, conspiracy rumors seem to be just a pernicious special case of legends and fables, examples of which pervade the popular culture of all races, mostly via advertising and the entertainment media, with themes—"Giants, dwarves, fairies, witches, mermaids, anthropomorphic objects, and personified principles"—similar to those of the fairy tales and peasant *Märchen* of yore (Dégh, 1994, p. 36).

***Survey Data Concerning Three Rumors of Conspiracy Against Blacks.*** Table 3 presents and analyzes results from a 1990 survey in which both Blacks and Whites indicated their belief in three specimens of such conspiracy rumors. According to these three rumors, the government targets officials for prosecution because they are Black, makes drugs available in Black neighborhoods to harm

**TABLE 3**  
**IQ Commensurability of Black-White Differences in Judging Credibility of Three Rumors, New York City Sample**

	Not replying "Almost certainly not true" (%)			Implied Critical IQ		
	Blacks	Whites	Difference	Blacks <sup>a</sup>	Whites <sup>b</sup>	Difference <sup>c</sup>
<b>Rumor 1</b>						
"Some people say the Government deliberately singles out and investigates Black elected officials in order to discredit them in a way it doesn't do with White officials. Do you think that is true, or that it might possibly be true, or that it is almost certainly not true?"	83.59	43.24	40.35	98.22	99.01	-.79
<b>Rumor 2</b>						
"Some people say the Government deliberately makes sure that drugs are easily available in poor Black neighborhoods in order to harm Black people. Do you think that is true, or that it might possibly be true, or that it is almost certainly not true?"	66.02	24.49	41.53	91.16	90.48	.68
<b>Rumor 3</b>						
"Some people say the virus which causes AIDS was deliberately created in a laboratory in order to infect Black people. Do you think that is true, or that it might possibly be true, or that it is almost certainly not true?"	36.33	9.29	27.04	81.64	80.10	1.54
<b>Algebraic mean of 3 surveyed rumors</b>						.48
<b>Absolute mean of 3 surveyed rumors</b>						1.00

*Note.* From a computer tally of response frequencies in the *New York Times/WCBS News Poll*. Race Relations in New York City, June 17-18, 1990 (DeParle, 1990). White  $n = 484$  ( $ME = 4$ ), Black  $n = 408$  ( $ME = 5$ ). For samples this size, the  $SE$  of the difference in mean IQ  $\approx .97$ . This means that 95% of the time, the IQ difference of the samples would be within  $\pm 1.94$  IQ points of its true value.

<sup>a</sup>Based on mean IQ = 86.0 and  $SD = 12.5$ . These parameter estimates apply to Northeastern Blacks, and are based on past research in Philadelphia, for the mean, and in PA, NY, and NJ and the Southeastern U.S., for the  $SD$ , as described elsewhere (Gordon, 1986, Table 2).

<sup>b</sup>Based on mean IQ = 101.8 and  $SD = 16.4$  from the 1937 Stanford-Binet normative sample (Terman & Merrill, 1960, Figure 4), used in Gordon (1976, 1986).

<sup>c</sup>The pattern of ascending B-W differences in conjunction with counting from the left tail of the IQ distribution suggests that the ratio of Black  $SD$  to White  $SD$  in this sample is underestimated by the values employed. If a larger IQ  $SD$  of 14.2 is substituted for Blacks, the three B-W differences in critical IQs lie close to the re-estimated  $SE = 1.02$  of the difference in mean IQ for samples of this size. An error in the difference between mean IQs, unlike one in the  $SD$  ratio, produces a constant error in the difference between critical IQs, and could account for a roughly constant observed difference, therefore, that is close to the  $SE$  in magnitude.

Blacks, and created AIDS to infect Blacks. (A reflowering of the rumor concerning drugs, in August 1996, is discussed in Appendix D.)

Before the results of testing the model are described, it should be noted that the response percentages in Table 3 reveal several important points that tend to be overlooked by folklorists. Specifically, opinion on the three rumors turns out not to be monolithic; some Blacks reject them. Folklorists offer no explanation of this bifurcation, tacitly taking some dissent for granted. The survey's inclusion of Whites reveals that such beliefs are not the exclusive property of the Black community, despite popular identification of that community as their host. Whites believe the rumors too, only less often. Rumor experts offer no account for why, if a rumor is believed by a certain percentage of Blacks, it will also be believed by a certain, but smaller, percentage of Whites. Ad hoc hypotheses based on simple ordering of the two races in sociological respects would probably suffice to quell curiosity. The rumors themselves are not equally believed within a race. Some are believed more often than others, and the rank ordering of belief turns out to be the same for Blacks and Whites, just as it is, with minor exceptions, for the difficulties of more numerous and closely spaced intelligence test items (e.g., Gordon, 1984, Table 14).

There is thus much evidence of structure in the results, but folklorists specializing in rumors are not trained to consider structure. Differences in degree of belief can be passed off as inevitable variation, and similar ordering in both races can be dismissed as expectable, post hoc. Exactly what function such beliefs concerning Blacks would perform in the White community is a question not raised. Presumably, it would have to be either a different function or the function performed by the beliefs for Blacks might have to be revised so as to apply to Whites. As we shall see, the population-IQ-outcome model accommodates both the within-group and between-group variations in belief.

More so than other measures considered thus far, survey data such as those in Table 3 can be viewed as measures of the *point* prevalence of an opinion, "which is [recall] the probability that an individual in a population will be a case [i.e., hold the opinion] at time  $t$ " (Kleinbaum et al., 1982, p. 118). This measure of prevalence is subject to analysis with the IQ model too. (Procedural decisions for dichotomizing survey response categories and designating, however tentatively, one of the two resulting answers "correct" and the other "incorrect" are detailed in Appendix D.)

**Results and Interpretations.** Despite the hazards of new forms of error peculiar to survey data (see Appendix D), applying the IQ model to the race-specific prevalence of the three rumors in Table 3 yields surprisingly good fits. The substantial differences between Blacks and Whites in these beliefs prove to be closely commensurate with differences in the IQ distributions of their populations. Taking the algebraic average of differences in critical IQ, which permits positive and

negative errors to cancel, produces a mean difference of only .48 IQ points, a virtually perfect fit. The small size of the average absolute difference in critical IQs, 1.0, indicates that the small algebraic mean does not conceal consistently large, but offsetting, errors in both directions.

Beliefs are commonly viewed as cultural phenomena, and culture, above all, is considered a socially constructed reality. It would appear from the results, however, that group differences in culture, perhaps unlike culture itself, are not necessarily themselves simply a matter of social construction, as they seem, in this case at least, to track population differences in IQ.

Oftentimes, the rumor data in Table 3 would be viewed as evidence of credulousness, although viewing conspiracy themes as evidence of credulousness mixes evident suspiciousness with supposed unsuspectingness. When someone marks a wrong answer on a multiple-choice test, we seldom speak of that individual as being "credulous"; we simply say that the individual failed to solve the problem and, perhaps, guessed wrong. The fit of the model in Table 3 argues for viewing supposed credulousness as just another form of failure in problem solving: in evaluating information, evaluating sources, integrating the whole, ironing out inconsistencies, and discounting personal feelings and other distractions as irrelevant.

Any problem is too difficult for some persons, and the three rumors, in varying degree, exemplify this familiar fact for the persons polled. As the IQ model fit them all, differences between the rumors in percentages of believing respondents can be understood simply as differences in their difficulty as problems at the time in question rather than as vagaries of fashion. Item 1 was the least preposterous rumor, and hence the hardest problem, because one had to contend with highly visible prosecutions of Black officials (the crack cocaine case against Washington, DC, Mayor Marion Barry was much in the news at the time of the survey) as well as with aspects of the justice system that are largely inaccessible to public view. Not everyone could be counted on to read responsible discussions of the issue of selective prosecution of Black officials in media, if such could be found (e.g., Cooper, 1996). Few would realize that the magnitude of the Black/White ratio of criminal prosecutions of elected officials (5.3:1) was about the same as that for adults in general (cf. line 10 of Table 1 and Davidson, 1996, p. 40). Hochschild (1995, pp. 106–107) too considered the items to be ranked in plausibility as they are numbered in Table 3. (See Appendix D for a discussion of results from Hochschild that, at first glance, seem to be inconsistent with those for IQ presented here.)

Table 3 is not the only example of conspiratorial opinions that fit the model. In the immediate aftermath of the later O.J. Simpson verdict, respondents to a national survey were asked whether they agree or disagree with the statement, "The white establishment is always trying to bring down successful black people" (Morin, 1995, p. A34). In content, this item is similar to the 1990 conspiracy

rumor items, although it is not couched as a rumor per se (i.e., "Some people say . . ."). Respondents agreed with the 1995 statement 52% of the time if Black ( $n = 312$ ), 12% if White ( $n = 315$ ), producing a large percentage difference (margin of error = 6 for each percentage). Critical IQs for this item were 84.5 and 82.5, a difference of only 2.0, which indicates a good fit despite the fact that the survey was conducted during a time of heightened sensitivity to such issues among many Blacks. The small observed difference in critical IQs could well be explained by sampling error. The percentages locate this item between Rumors 2 and 3 in difficulty, incidentally, which seems reasonable on the basis of its content.

### Surveys of Everyday Opinion Concerning the Simpson Case

***Background and Motivation for Examining the O.J. Simpson Surveys.*** The not guilty verdict on October 3, 1995, of the nearly all-Black jury in the 9-month, \$9 million Simpson double-murder trial after less than 4 hr of deliberation, along with nationally televised images of Blacks from all walks of life rejoicing over the announcement, left much of the nation stunned. A degree of racial polarization had been quantified by many months of polling and thrust before the nation that few persons, Black or White, viewed with complacency, especially now that it had been made palpable by a verdict in what media had dubbed "the trial of the century" (e.g., Bollinger & Hoffmann, 1995; Kramon, 1995; Reibstein, 1995).

Black columnist Carl Rowan (1995, p. 15A) earlier had complained of the polls, "Their constant harping on differences in racial attitudes creates the potential for stupid racial strife." Defense attorney Robert Shapiro (1996, p. 355) later blamed the polls for splitting the Simpson case "down racial lines," although it has since been reported that it was Shapiro himself who first conceived of a Simpson defense strategy that would be based on an alleged racist conspiracy (Cochran, 1996; Toobin, 1996). In any case, the polls did not invent the race differences, they only reported them. If the verdict made the polling data palpable, the abundant polling data in return made it impossible to dismiss the jury verdict as just an aberration of small sample size that held few implications for the wider democracy.

President Clinton expressed surprise at "the depth of the divergence" (Harris, 1995, p. A13). All across the political spectrum others used phrases like "chasm" (Hallow, 1995, p. A8), "gaping racial divide" (Eisner, 1995, p. C7), "two separate societies" (Dowd, 1995, p. A29), "two nations—one black, one white" (Greenberg, 1995, p. A20; King, 1995, p. A29; Thernstrom, 1995), and "[a lesson in] how differently blacks and whites see the world" (Littwin, 1995, p. 5E; Cohen, 1995). No one sharing concern over a situation that would evoke such responses can afford to ignore opportunities to understand better what had occurred.

Initial impressions among many Whites were that the jury had responded to

emotional racial appeals used by defense attorney Johnnie Cochran in his closing argument, “Maybe you’re the right people . . . to say: ‘No more!’” (variously punctuated in Cochran, 1996, p. 349; Cohen, 1995, p. 25; Deutsch, 1995; Safire, 1995, p. A23), and to being told by Cochran that they were “the ones to send the message” on behalf of all Blacks (Littwin, 1995; Reibstein, 1995, p. 28; Toobin, 1996, p. 421). Cochran, practiced in employing the racial conspiracy tactic to good effect on behalf of prior clients (Toobin, 1996, pp. 180–183, 381–382), not without justification when cases involved police brutality (Cochran, 1996), had catalyzed a police conspiracy theory many regarded as having been concocted by the defense (Bollinger & Hoffmann, 1995; Bugliosi, 1996; Margolick, 1995; Thernstrom, 1995; Toobin, 1996). Suspicions about the defense theory were dramatically reinforced soon after the verdict when fellow defense attorney Shapiro complained—however inconsistently in view of his own role in initiating the race strategy that Cochran then executed with virtuosity (Toobin, 1996, pp. 154, 438–439)—“Not only did we play the race card, we dealt it from the bottom of the deck” (Achenbach, 1995, p. A27; Shapiro, 1996, p. 349). These suspicions persist in many quarters still (e.g., Darden, 1996; Toobin, 1996).

A misunderstood farewell salute to Simpson from a juror who is a former Black Panther evoked memories of the “Black Power” movement for some observers (Adams, 1995; Dowd, 1995; Safire, 1995; Thernstrom, 1995). The juror later explained that his use of the gesture, commonly employed with nuanced meanings among Blacks, was meant only to convey good wishes to the newly freed defendant (“Juror #6,” 1995). The verdict was often viewed as an instance of jury nullification (viewed as the determination to ignore evidence of guilt in a particular case, but originally a rejection of a law considered unjust) and an expression of racial anger over police injustices (Darden, 1996; Goldberg, 1996; Holden, Cohen, & de Lisser, 1995; Hopkins, 1995; Kennedy, 1994; Smith, 1995), possibly justified, according to one Black law professor, by the high incarceration rate of Black males (Butler, 1995a). Such special motives, if true, would certainly account for the outcome. Motives, however, can always be adduced to account for any human behavior, *ad hoc*.

Other commentators ascribed the verdict simply to different racial perceptions of police, based on different experiences at their hands and instances of unprofessional police behavior (Butler, 1995b; Cannon, 1995). One law professor invoked differences in knowledge, arguing that Blacks were “better informed” about the justice system than Whites (Duke, 1995, p. A22), which might be true in some respects yet not in others. A social constructivist media critic who commented on the trial and its forensic evidence distinguished “Blackstream Knowledge” from mainstream or White knowledge, and contended that “some African Americans believe that scientific truth is not universal but white, if not in its essence, at least in its uses” (Fiske, 1996, pp. 191, 264). In early mock juries conducted for the

prosecution, Blacks had, in fact, remained largely unmoved by hypothetical scientific evidence that would establish Simpson's guilt beyond question (Toobin, 1996, p. 191).

Regardless of doubts, some persons thought it important to maintain public confidence in the acquittal, so much so that a bar association award to the prosecuting attorneys was blocked because it would be a slap in the face to the Simpson jury ("Clark, Darden," 1995). A noted journalist severed the issue of truth from the fact of difference, regarding as "almost irrelevant [whether] the black or white judgment might be more nearly correct," and holding that "what mattered was the demonstration that whites and blacks . . . see themselves in different worlds" (Wicker, 1996, pp. x-xi). But the reconciliation of differences without recourse to truth would compromise endlessly the position of the factually supported side.

When the nine Black and three other jurors began to explain themselves, they denied race played a role, and emphasized that reasonable doubts brought about their verdict (Adams, 1995; Egan, 1995; "Hispanic Juror," 1995). Their sincerity, feelings of conscientiousness, and lack of apparent animosity toward Whites were plain, especially in televised interviews, and drew support from many commentators, Black and White (Hentoff, 1995; Kramon, 1995; Raspberry, 1995a, 1995b).

Some Black commentators, while defending the jury or, earlier, the possibility of a not guilty verdict, took pains to inform the nation that not all Blacks thought Simpson innocent: "Let me say straight out, the prosecution convinced *me*," wrote columnist William Raspberry (1995b, p. A23). Another columnist, Chuck Stone, stated, "I think O.J. is guilty. All the evidence points to it" (Sachs, 1995, p. 37). Columnist Carl Rowan (1995, p. 15A) used a division within his own family to demonstrate that there was "no racial monolith on O.J.," a fact well documented by the opinion polls, but there was an enormous race difference, and the custom of reporting just one of the two percentages involved in a dichotomy left what split in opinion there was among Blacks less salient, as well as totally unexplained. Why did a majority of Blacks feel one way, a substantial minority a different way (including uncertainty)?

Once the jurors began speaking out, analyses began to shift away from attitudinal and political explanations. The issue of intelligence was raised, often using the crude terms favored in media that seem to discredit the very idea of intelligence: "Are black people racist or just stupid?" (Butler, 1995b, p. C1; Hentoff, 1995; Milloy, 1995). A few astute commentators, however, saw the verdict, what ought to be a product of informed judgment and subjective estimation of probabilities, as an inadequate analysis of complex evidence presented by over 100 witnesses in 35 to 45 thousand pages of transcript (Goldberg, 1996, p. 377; Toobin, 1996, p. 375) to a jury whose educational level had been reduced in the course of jury selection with the help of peremptory and other challenges (Dow, 1995; Rothwax, 1996, pp. 228-229), and, especially, by exclusions based on the hard-

ship of a long trial to high-income individuals who would lose pay (Toobin, 1996, p. 197).

The key intellectual problem, as one judge saw it, was that the jurors “were remarkably poor evaluators of the facts” and had consistently confused logical possibilities with grounds for reasonable doubt by neglecting to give sensible consideration to probabilities (Rothwax, 1996, pp. 226, 229). A law professor cast the problem more broadly: “Modern trials hinge on complicated assessments of economic models and DNA evidence . . . Lay juries are simply not equipped to perform this assessment” (Dow, 1995, p. A32). Coincidentally, a national survey of adults in 1995 had found that “only one in five Americans [21%] can provide a minimally acceptable definition of DNA” (National Science Board, 1996, pp. 7–8, Appendix Table 7-7). During expert but often dull DNA testimony, the jurors were visibly bored (Shapiro, 1996, p. 353; Toobin, 1996, p. 345). DNA was mentioned 10,000 times (“Simpson Trial & Trivia,” 1995). When presented with the statement, “The Simpson jury just wasn’t smart enough to understand the evidence in the case,” 26% of Whites, but also 10% of Blacks, agreed (Morin, 1995, p. A34). If it is reasonable to regard the task of jury service as a kind of job, validity generalization theory would indicate that performance of that job must be related to *g*, most especially in complex trials (e.g., Gottfredson, 1986b).

Of interest here is the fact that when the CBS-TV News program *60 Minutes* investigated an allegation of jury tampering 6 months after the verdict, the juror whose dismissal well into the prosecution’s case was being blamed on what definitely appeared to be a fabricated letter turned out to be an extremely articulate and self-possessed White woman, who, the defense’s own jury consultant conceded, had been sized up as a person with “leadership qualities” (Rosenberg, 1996). This woman, “one of the best-educated members of the jury” (Rosenberg, 1996) and the only one to volunteer what DNA stood for during jury selection (“deoxyribonucleic acid”), has stated that she became convinced Simpson was guilty and that, if necessary, she would have hung the jury rather than vote for acquittal in view of the evidence (see also Darden, 1996, pp. 293–295; Noble, 1996; Shapiro, 1996, pp. 285–286; Toobin, 1996). The issue of how smart the jury had to be to win an acquittal was evidently not lost on whomever was responsible for the mysterious letter, which alleged in a now highly suspect manner that the woman and her husband were negotiating a book contract, and that the book would describe why the author hung the jury in a trial that was far from over.

Especially lacking in the final Simpson jury was an appreciation for the multiplication of independent probabilities needed to account for the full weight of incriminating evidence (Paulos, 1995), what former Los Angeles Police Chief Daryl Gates called, “the extraordinary number of simple common sense coincidences” (Norbom, 1995, p. 4). Bugliosi (1996, pp. 218–219), making the same point, required two pages just to list them, and Goldberg (1996) enumerated 55 such items. Even defense attorney Alan Dershowitz (1996, p. 108) was guilty of

faulty probability reasoning when he correctly pointed out that fewer than 1 in 1000 wives who are abused by their spouses, as Mrs. Simpson had been, are later killed by them. One dismissed juror similarly asked, "What does domestic violence have to do with murder?" (Darden, 1996, p. 332; Rothwax, 1996, p. 230), and a final juror stated she had no use for an argument based on wife abuse (Streisand, 1995, p. 38; see also Toobin, 1996, p. 191). A mathematician replied to Dershowitz, using the relevant conditional probabilities, that "if a man abuses his wife and she is later murdered, the batterer is the murderer more than 80 percent of the time" (Paulos, 1995, p. C7).

Without knowing that the probability was over 80%, as few persons would, one could have some idea that batterers of murdered wives are often implicated in the homicide simply from reading newspapers, a source, recall, that 27% of adults do not consult with any great regularity (Barton & Jenkins, 1995, Table 3.6). Not one of the final Simpson jurors read a newspaper regularly, and eight did not watch the evening news on television (Bugliosi, 1996, p. 96). This was a jury that could have benefited from the two-step flow of communication if only it had included some well-informed persons.

In the wake of dismay over the verdict, information concerning suspiciously similar jury behavior in other, less publicized cases began to come forth (Haber- man, 1995), showing that the surprising outcome of the Simpson trial was by no means an isolated phenomenon. In the Bronx, Black defendants in felony cases were acquitted 47.6% of the time by predominantly Black and Hispanic juries, almost three times the national rate for all defendants, and elevated acquittal rates were reported from other heavily Black venues: Washington, DC, and Wayne County, Michigan (Holden et al., 1995; Purnick, 1995). In California, about 10% of criminal cases resulted in hung juries (Janes, 1995; Riley, 1995), a statistic that ought to temper suggestions that mixing citizens from the mostly White (and entirely non-Black) Simi Valley jury that acquitted officers in the Rodney King beating and the mostly Black Simpson jury would necessarily have ironed out misperceptions in both panels, each of which denied that its verdict was based on race (Cannon, 1995).

In Baltimore, a jury with 11 Black members gained national attention by returning a not guilty verdict for a Black defendant in the murder of a Korean student despite testimony from four eyewitnesses and two others told by him that he had committed the crime (Holden et al., 1995). When the Korean community protested that the acquittal had been racially motivated, these jurors too defended their verdict on the basis of reasonable doubt: It was possible the witnesses had lied, been mistaken, or were actually the guilty parties themselves. The jurors denied racial motivation, sincerely in my opinion, but their explanation, as has been said of writings by Lewis Carroll, was not "illogical so much as unconstrained by common sense" (T. Lewis, 1995, p. A18). They had invented their own highly speculative theory of the crime. Korean American distrust of the out-

come was not diminished when news was released that, prior to the verdict but unbeknownst to the jury, the defendant had offered to accept a 40-year sentence in a plea bargain, which the victim's family had rejected as too lenient (Heard, 1995; Hermann, 1995; James & West, 1995).

As in the case of police, now and then the behavior of some juries or single Black jurors, sometimes with help from misunderstandings by others, lent credence to the idea that racism was deeply implicated everywhere (Haberman, 1995). In Maryland, one of six Black jurors, claiming to be influenced by the conspiracy defense in the Simpson case, hung his 12-person jury on the theory that the defendant's finger prints should have been recovered from a tossed-away handgun; police later stated that it is sometimes difficult to lift prints from pistol handgrips (Jeter, 1995). In an attempt to provide balance, some commentators noted that Black jurors do convict Black defendants routinely, glaring exceptions and atypical statistics notwithstanding (Broder, 1995; Goldberg, 1996, p. 354).

In what might be considered the second act of the postverdict Simpson drama, calls for a review of the American jury system began to be heard. Polls revealed a decrease of confidence in the racial impartiality of jurors among Whites (but an increase in confidence among Blacks) that occurred between the time of the verdict and 4 months prior (Moore & Saad, 1995, p. 8). Some experts suggested abolishing the jury system and employing a panel of judges, the practice in many European nations (Dow, 1995; Riley, 1995). Others favored abandoning the requirement of unanimity (Amar & Amar, 1996; "False Lessons," 1995; Goldberg, 1996, p. 365; Riley, 1995; Rothwax, 1995, pp. 211–215), a policy that might be perceived, especially by Blacks, as racially motivated itself, as it would make convictions more likely. Such a change could fuel Rumor 1 (concerning selective prosecution of Black officials) in Table 3, for example. When and where reverberations from the Simpson trial will end, no one can tell.

What did account for the more than 40-point race split in public opinion over Simpson's guilt? Was it a matter of problem-solving ability, as in the case of the three rumors, or was it something that had little to do with intelligence and more to do with social psychology that might even have had the opposite relation to intelligence within groups? In particular, being able to rule out racial ill-will or prejudice as a factor in its own right by replacing that hypothesis with a more adequate one might help defuse antagonism. If the split in public opinion could be understood in terms of intelligence, the verdict itself might become more comprehensible, as it is reasonable to assume that the jury represented a selection from the wider population, but one biased against the inclusion of brighter individuals ("A Guide," 1995; Amar & Amar, 1996; Dow, 1995; Goldberg, 1996, p. 351; Rothwax, 1996). Although that in itself might not be a welcome conclusion either, understanding creates a basis for more constructive policy.

As an explanation of the verdict, the race factor hypothesis continues to linger. It has been reactivated by Toobin, for example, in what I consider to be one of the

three most informative books on the Simpson trial (the others are Bugliosi, 1996, and Goldberg, 1996). Toobin, who is convinced Simpson is guilty by “any rational analysis of the . . . evidence” (1996, p. 435), pondered the outcome and what he was able to learn about the actual deliberation. Toobin concluded that race had in fact been the overriding consideration affecting the verdict, despite the denials of jurors, but his own words belie that conclusion.

All the black jurors denied that race played any role at all in their deliberations or their decision. To me, this is implausible. The perfunctory review of nine months’ worth of evidence; the focus on tangential, if not actually irrelevant, parts of that evidence; the simply incorrect view of other evidence; and the constant focus on racial issues both inside and outside the courtroom [i.e., in their sequestered living arrangement as well]—all these factors lead me to conclude that race played a far larger role in the verdict than the jurors conceded. (Toobin, 1996, p. 437)

Ironically, there appear ample signs of intellectual overload in the very description of the jurors’ performance that Toobin provides to support his race diagnosis: “perfunctory review,” misplaced attention, “incorrect” views of evidence, and overreliance on a simplistic race hypothesis both on the job as jurors and in their sometimes prickly interpersonal relations. As an explanation, race is typically underanalyzed and overworked, and its role in this instance may not be the simple one of racial solidarity in defiance of the facts that Toobin appears to support. At an earlier point in his book, Toobin (1996, p. 11) noted that the defense’s counter-narrative of racial conspiracy “needed a receptive audience,” which it would find among Blacks in the jury pool. Toobin failed to consider why the race issue proved so effective, why the audience in question would be “receptive,” aside from their race. That polls clearly revealed the race issue had not totally overpowered a substantial minority of Blacks seems to have posed little problem for Toobin; such individuals were implicitly relegated by him to the status of unexplained racial exceptions. In his treatment of the verdict, Toobin (1996, p. 169) seemed to be honoring the same taboo against drawing “attention to the intellectual limitations of any African-American” that he ascribes to mainstream journalists.

Against such a background, prior success with the rumor data in Table 3 suggested that surveys of opinion on the Simpson case, if tested with the IQ model, might be able to shed light on key questions concerning the verdict. If the population–IQ–outcome model should fit the polling data, that would support the hypothesis that intelligence differences lay behind the opinion differences. When initial tests on one or two published polls, scribbled in the margins of newspapers, proved suggestive (in view of sampling and other potential sources of error variation), a decision was made to seek out data from all identifiable Simpson polls in order to gain reliability and greater representation.

*The Search for O.J. Simpson Surveys.* The Nexis data base was searched for media text linking “Simpson” and “polls” and related words so as to identi-

fy organizations that had conducted surveys on the Simpson case. Copies of their poll reports were requested, and items suitable for analysis were identified. Later requests targeted data pertaining only to such items. Several organizations sent only their more recent reports, but often these referred to earlier polls, which were then specifically requested, and the process repeated, sometimes more than once.

At a later point, the Roper Center's Public Opinion Online, which serves as an archive for polling organizations, was searched for additional surveys, which are filed by item, under the keyword "Simpson." The contents of all items were scrutinized to discover any additional ones of possible interest as well as unprovided reports from other sources. When checked for the last time (on November 13, 1995), Public Opinion Online contained 647 Simpson items on various topics, all reported in summary form only, that is, without analyses by race. Omitted polls newly uncovered were again requested specifically from the organizations that had responded earlier. Bivariate survey results by race that still remained unavailable from the original sources were analyzed and provided by the Roper Center, for a fee. One local poll came to attention through television. A few overlooked polls were discovered as late as March 1996 and were incorporated in the analyses, which include some polls conducted just after the verdict.

Two forms of survey item were identified early on as being of main interest and then pursued exhaustively. One dealt with belief in Simpson's guilt or with agreement with the verdict. The other dealt with whether a police conspiracy or Detective Mark Fuhrman, a prosecution witness depicted as having strong anti-Black sentiments, had attempted to frame Simpson (for balanced accounts of the Fuhrman issue following investigation of him after the trial, see Bugliosi, 1996, pp. 129–135, 261–262, 333–334, Butterfield, 1996, Goldberg, 1996, and Ross, 1996). There were also two methodological variants, trichotomous, with three responses including "don't know" (DK) and "no answer" (NA) as one, and polytomous or multicategorical items providing for degrees of agreement, such as *definitely guilty*, *probably guilty*, *probably not guilty*, and *definitely not guilty*. Both variants were dichotomized (see Appendix D). The resulting four combinations of content and method, each containing minor variations, have been analyzed separately. Data from 86 identified survey items meeting these specifications were compiled.

Because a DK or NA response is more defensible to a question about a particular fact not a matter of public knowledge, such as whether Simpson committed the crime, than to a generally known fact, such as whether the moon is made of green cheese, DK and NA are now treated as not-wrong answers (see Appendix D). For purposes of analysis, beliefs in Simpson's innocence and in a conspiracy to frame him were cast in the role of wrong answers. The White samples are ample enough by survey standards, ranging in size from about 300 to about 1000, although minuscule in comparison with the size of samples on crime, parenthood, and HIV in Table 1. Black samples, however, were often startlingly small, whenever

Blacks were not oversampled so as to achieve an  $n$  of about 300 to 400. The smallest Black sample reported, for example, is only 66.

*Applying the Population-IQ-Outcome Model to the Simpson Survey Data.* Table 4 summarizes results of testing the IQ model on each of the four combinations of item from surveys conducted between June 22, 1994, and October 5 through 7, 1995 (space limitations preclude reporting critical IQs for all 86 items). Summary percentages of aggregated responses are shown in Table 5.

The first entry in Table 4 is for trichotomous "not guilty" items. Although there were clearly instances among the many individual items of less than impressive fit, which again begins at differences in critical IQ of  $\pm 3.0$ , in many cases the model fit well. Just over two thirds of the 42 items exhibited differences in critical IQ that were smaller than 3.0 IQ points. Although differences in critical IQs were sometimes as large as 5 and 7, the average algebraic fit, which often gives sampling errors in both directions a chance to cancel out, turns out to be excellent, at 1.0. (The average absolute fit is 2.2 IQ points.)

Turning to the 28 polytomous "not guilty" items, there were some extremely poor fits. Differences in critical IQs reached as high as 7 and 9. However, the model again fit well in many cases, with 75% of the differences under  $\pm 3.0$ . The

**TABLE 4**  
**IQ Commensurability of Responses to Survey Items Concerning Guilt and Conspiracy in O.J. Simpson Case: Mean Algebraic and Mean Absolute B-W Differences in Critical IQs**

Response	Algebraic	Absolute
<b>Simpson guilty? (trichotomous), 44 survey items</b>		
Not guilty	1.0	2.2
<b>Simpson guilty? (polytomous), 28 survey items</b>		
Net not guilty	1.6	2.6
<b>Conspiracy/frame by planting glove? (trichotomous), 10 survey items</b>		
Yes/planted	-.2	1.8
<b>Conspiracy/frame by planting glove? (polytomous), 4 survey items</b>		
Net likely/planted	-.8	.8
<b>Grand mean, 86 survey items weighted equally</b>		
Not guilty, conspiracy	.98	2.24

*Note.* Surveys from and conducted for the following organizations are included: *ABC News*; *ABC News/Washington Post Polls*; *Associated Press Polls*; *Buller Poll* for *WBAL Channel 11 News*, Baltimore; *CBS News Poll*; *The Field Poll*, *The Gallup Poll* for *CNN/USA Today*, analyzed and provided by *The Roper Center*, University of Connecticut; *The Harris Poll*; *Los Angeles Times Polls*; *NBC News and Dateline NBC Polls*; *NBC News/Wall Street Journal Polls*; *Newsweek Polls*; *Washington Post Polls*; *Yankelovich Partners for Time*. The contributions and assistance of these organizations are gratefully acknowledged.

average algebraic difference is again small, at 1.6. (The average absolute fit is 2.6 points.)

The results of testing the model on 10 trichotomous items concerning the police conspiracy theme appear next in Table 4. The poorest single fit was only as large as 4.4, and the average algebraic fit is virtually perfect, at  $-.2$ . Critical IQ differences were less than  $\pm 3.0$  in 90% of the cases. (The average absolute fit is 1.8 points.)

The fourth entry in Table 4 contains results for four polytomous items on the conspiracy theme. Because all four differences between critical IQs were negative in sign, the average absolute difference in the case of these few items and the average algebraic difference are equal in unsigned magnitude. There was no opportunity for positive and negative errors to cancel. Nevertheless, all four items show acceptable or excellent fit to the IQ model, and the average algebraic (and absolute) difference in critical IQs is only  $-.8$  (and  $.8$ ).

The final entry in Table 4 is the mean algebraic difference in critical IQs for all 86 items,  $.98$ . Such a low value leaves no doubt that the fit of the population-IQ-outcome to the Simpson survey data is excellent. (The average absolute fit is 2.24 points.)

That the Simpson items overall may be somewhat heterogeneous in their true levels of response is suggested by differences in mean percentages (displayed in Table 5) between combinations of item (e.g., a range from 60.42% to 73.98% for Blacks) and by differences among the numerous individual items themselves within combinations (not shown). There are also the indicated differences in item content. The heterogeneity is not nearly as great as in the case of the rumor data in Table 3, however. The *SD* of the three Black and three White percentages in Table 3 are 23.9 and 17.0, respectively, whereas for the 78 Simpson items to which national IQ parameters apply, the corresponding *SD*s of percentages are only 12.3 and 8.1. Eight items from California surveys have now been excluded from this form of averaging for simplicity, because they employed slightly different IQ parameters more appropriate for that state (from Mercer, 1984, Table 9), a fact which would itself be a minor source of heterogeneity.

Squaring the *SD*s to obtain the variances, the Simpson data were only 26% and 23% as variable as the rumor data. In view of this much-reduced variation, it was decided to explore the effect of averaging the percentages before applying the IQ model. The smaller amount of heterogeneity in levels of percentages would dampen the disruptive effect of nonlinearity in the transformation of percentages to normal deviates required by the model and so render the procedure less questionable than would have been the case for Table 3.

Table 5 presents results from averaging the percentages within race before applying the IQ model, when items are weighted equally (sample sizes are not always known). Separate analyses are reported for each of the four combinations of response type and content. One of the combinations now produces a virtually

**TABLE 5**  
**IQ Commensurability With Mean Responses Concerning Guilt and Conspiracy**  
**in O.J. Simpson Surveys<sup>a</sup> and in NYC Tawana Brawley Survey<sup>b</sup>**

Response	Mean Percentage		Critical IQ		Critical IQ Difference
	Blacks	Whites	Blacks	Whites	B - W
<b>Simpson Surveys</b>					
Simpson guilty? (trichotomous), 42 survey items					
Not guilty	61.42	19.46	87.60	87.68	-.08
Simpson guilty? (polytomous), 24 survey items					
Net not guilty	61.94	16.64	87.78	85.92	1.86
Conspiracy/frame? (trichotomous), 8 survey items					
Yes/planted	68.38	26.00	90.06	91.25	-1.19
Conspiracy/frame? (polytomous), 4 survey items					
Net likely/planted	73.98	30.23	92.21	93.31	-1.10
Grand mean percentages, 78 survey items weighted equally					
Not guilty, conspiracy	62.94	19.82	88.12	87.89	.23
<b>Brawley Survey<sup>c</sup></b>					
Any response except lied	67	27	91.5	91.8	-.3

*Note.* <sup>a</sup>National IQ parameters employed for all O.J. Simpson analyses (see Table 2). Simpson results include one local survey in Baltimore, which was assigned national parameters because Maryland is a border state (i.e., between Northeast and South regions). Eight California and Los Angeles surveys have been omitted.

<sup>b</sup>Survey data from Edsall and Edsall (1992, pp. 2239-240); analysis of the New York City survey on Brawley employs IQ parameters for Northeast, as for Philadelphia (see Table 2).

<sup>c</sup>New Yorkers not saying Tawana Brawley lied after 1988 grand jury determined her story had been fabricated.

perfect fit, and the other three have quite acceptable fits. There is little difference in fit according to type of item content (guilt vs. conspiracy).

The four differences in critical IQs based on the mean percentage method in Table 5 can be compared with four algebraic mean differences in critical IQs (supplied here in parentheses) for the exact same 78 national survey items. Reading the former from the top down in Table 5, these are -.08 (.92), 1.86 (2.10), -1.19 (-.98), and -1.10 (-.80). The correlation between absolute magnitudes of the above data for the two methods is .71, and the means of those magnitudes are 1.06 and (1.20), indicating that both methods tended to place the four results in about the same order, and neither appears much superior to the other.

In view of the successful fit between the model and the four combinations of Simpson items in Table 5, remaining cautions over heterogeneity of true response levels in these data were set aside, and the model was applied to the average

percentages for all 78 items to which the national IQ parameters in Table 2 were applicable. This meant mixing together in one analysis items from the four combinations of content and method so that whatever heterogeneity might be peculiar to distinctions among those categories would no longer be isolated within separate analyses.

Even knowing the prior results, the minuscule difference in critical IQs for the 78 items, .23 IQ points (midsection of Table 5), is nothing short of astonishing. Going into the analysis of the untidy-seeming data from the many Simpson surveys, conducted at different points in the history of the trial by different polling organizations using different questions that were often preceded by different kinds of questions, and based sometimes on relatively modest samples of different sizes, who would have guessed that an overall fit to the intelligence model would emerge that was virtually perfect? Restated in the metric of percentages, the fit is so nearly perfect, no matter which method of aggregating results is used, that the aggregate percentage of "wrong" responses from each race could be predicted within less than one percentage point simply by knowing the percentage of such responses in the other race and applying the model. One could not ask for more compelling testimony to the power of aggregate data and the role of intelligence in everyday life.

The so-called and much-discussed Great Divide (Littwin, 1995, p. 1E) between Blacks and Whites in two major categories of opinion concerning the Simpson trial is explainable entirely on the basis of differences in the Black and White IQ distributions. The split between races over Simpson's guilt that supposedly reflected what social constructivists term "local . . . knowledge" (Fiske, 1996, pp. 267–268) turns out to be located far more precisely, when *g* and matching critical IQs are considered, than was anticipated by their broad conception of racially defined subcultures. Ad hoc explanations of outcomes based on supposedly specific knowledge may proliferate, but in this instance they cannot compete with IQ when tests more challenging than that of superficial plausibility are employed.

### **Everyday Opinion Concerning the Tawana Brawley Affair**

The Tawana Brawley affair lends itself to a test for the same population-level effects of intelligence that were wrongly attributed to race in the O.J. Simpson case. The details of the affair itself also illustrate how contextual effects are implicated in individual behavior. Once again, an example is chosen because of the availability of relevant survey data, its extensive publicity (which made its facts known to most citizens), and its impact on race relations.

Between late November of 1987 and early October of 1988, a case involving a Black teenager, Tawana Brawley, roiled emotions in New York City and State and the rest of the nation. Brawley and her mother, and soon her two Black lawyers and other activist advisers, claimed that she had been abducted and raped repeat-

edly by a gang of White racists, one of whom displayed a police badge, who then left her in a humiliating state, smeared with filth and marked with a racial epithet and “KKK.” Although the teenager, her mother, and their supporters refused to cooperate with the investigation in any way, some of them eventually went so far as to name an assistant district attorney publicly as one of the assailants. When evidence to support the story could not be uncovered, criminal justice investigators were subjected to a public campaign of extreme personal vilification, and allegations of an official conspiracy to cover up the crime, not to mention of actual involvement in the supposed rape, streamed forth virtually without interruption from the Brawley camp (“Evidence Points,” 1988; Taylor, 1992, pp. 79–80). For 1988, the *New York Times Index* lists 118 days on which articles appeared under “Brawley.” Thus, the Brawley case dominated the news, much as the Simpson case was to do 7 years later when, with 26.8 hr of network news time (“Simpson Trial,” 1995), it became the top story of the year.

Finally, a 7-month grand jury investigation concluded that the entire affair was a hoax, originally contrived by Brawley and her mother to cover up the girl’s 4-day absence from home because she feared punishment from her mother’s live-in companion (a wife-killer with a long history of violence) for skipping school to visit a jailed boyfriend and staying out until 5 a.m. (McFadden, 1988b). The poor judgment of a teenage girl, called forth in response to her immediate social context, had been highly amplified by her still wider social context to produce a costly cause célèbre, drawing its inspiration from the long tradition of racial conspiracy rumors. There was eyewitness testimony that she had placed herself in the garbage bag in which she had been found, as well as physical evidence linking her to the hiding place where she had spent the 4 days during which she was missing.

Like the Simpson case, the Brawley matter produced a sharp division in opinion between Blacks and Whites, again attended by justifiable concern on the part of some members of both races that irreparable damage was being done to race relations (“After the Grand Jury,” 1988; McFadden, 1988a). The final entry in Table 5 concerns the racial division in opinion that persisted in the aftermath of the grand jury’s report, as reflected in responses to a single survey asking whether the girl had lied. The Black–White difference in opinion concerning her veracity was almost perfectly commensurate with IQ, producing a critical IQ difference of only  $-.3$ . In the context of the rumor and Simpson findings, this lone datum now stands more secure.

### **How Do We Know Opinion Differences That the Model Fits Are Because of $g$ ?**

The consistently good fit of the IQ model to the four Simpson item combinations and the final excellent fit overall in Tables 4 and 5 imply the existence of an underlying structure that tends to shape opinions of the sort represented in Tables 3 through 5 so as to produce IQ commensurability for Blacks and Whites. Such a

structure is not, of course, necessitated, as all instances of good fit could result from what Bailey (1989, p. 70) called, when there existed only four tests of the model with IQ, "fortuitous coincidence." However, 125 trials of the model with SES variables in place of IQ failed to produce a single fit as good as those achieved by IQ for delinquency outcomes (Gordon, 1987). Even from parameters based on correlates of IQ, therefore, a good fit is not to be had for the asking.

Although the possibility of coincidence may be remote, the consistency between the implied structure and the Black and White IQ distributions does not prove that IQ is responsible for that structure. Any other normally distributed measured variable and its corresponding normally distributed latent trait having the same relative relations among means and *SDs* (or relations that fall within a limited range of trade-offs) could account for the outcomes. No other variable meeting these specifications, namely, that it mimic the two IQ distributions, seems to be available as a candidate, however.

Another candidate would not only have to shape the several kinds of opinion, it would also have to account for the delinquency, crime, single motherhood, HIV, and poverty prevalences in Tables 1 and 2 over the wide range of years involved, unless it is supposed that a different candidate exists for each type of outcome. SES, including years of schooling completed, has already been tried out in the candidate role for the delinquency data, and has failed (Gordon, 1987). Analyses of the failures led to new insights into the SES data, as the adequacy of fit depended on the degree to which SES distributions of individual differences served as surrogates for IQ distributions of individual differences.

Responding to my comparison between SES and IQ (Gordon, 1987) as sources for IQ commensurability in the first two delinquency studies to which the model was applied (lines 1–4 in Tables 1 and 2), Bailey (1989, p. 67), while agreeing that "IQ-commensurability remains an interesting phenomenon in search of an explanation," drew on applications in medical genetics to offer instead a "multifactorial threshold model of delinquency." Bailey's model of between-group differences in delinquency would include IQ as only one among many unspecified other variables, among which would implicitly be included contextual variables that I had identified as mediators of population IQ effects, but which he seemed not to count as a key part of the argument for an IQ model.

Being specific to delinquency, Bailey's model is less suitable for accounting for outcomes involving the heterogeneous variables that have now been examined here, and, no doubt, for whatever entirely new ones may yet be uncovered (on specificity arguments, see Jensen, 1984). One could broaden the hypothetical multifactorial variable that Bailey (1989, p. 67) called "the *lability* to delinquency" to encompass all such outcomes, but that would make its actual nature more mysterious yet if it is not to be conceived of as the total population effect of IQ or *g*. Or, one could tailor the hypothetical variable so as to include IQ with multifactors peculiar to each outcome, but that would require all the different manifesta-

tions of lability to possess about the same race-specific means and *SDs*, which strains credulity.

Bailey insufficiently credited the population–IQ–outcome model itself with acknowledging multifactoriality within groups, where the model is only probabilistic at each IQ level. Thus, in addition to IQ, individual outcomes are left to be determined by factors unrelated to both individual and population IQ (as witnessed in part by modest within-group correlations with proband's IQ). A model, recall, can be multifactorial within groups and still be effectively unifactorial between groups, even if the unifactor has multiple mediators in the form of its individual-level and contextual effects. The insistence that multifactorial models apply between groups as well as within groups would be a dogmatic one that is not always required by the data. Support for this realization in the intelligence domain is potentially the most important contribution of findings produced by the population–IQ–outcome perspective. The challenge for critics of the IQ hypothesis is to produce a viable alternative variable that can be named and measured independently of prevalences and that will also account for Black–White differences in those prevalences.

Ruling out confounding variables is usually the most troublesome aspect of causal modeling. But the view of experienced modelers is that if we are told only “that confounding variables may exist without any specific clues as to their identity,” we may as well proceed with the variables in hand (Asher, 1976, p. 12). Another important principle in deciding causality is that effects should be proportional to their causes (Einhorn & Hogarth, 1986). Unless the putative cause has been specified and measured, there is no way to apply this principle to alternative variables in the same manner it was applied to IQ. The feature of unspecified alternative variables that seems to make them attractive is that they can be asserted not to be IQ or *g*.

In searching for a concrete rather than hypothetical alternative to IQ, one could consider school achievement as a candidate, if one is willing to suppose, for example, that the Simpson results are the product of education instead of IQ. However, this would leave open which of the many possible measures of achievement to employ in the model. Years of schooling would not work, because the median gap between Blacks and Whites for recent young-adult cohorts is less than .12 White *SD* and shrinking (U.S. Bureau of the Census, 1979, Table 71; 1992, Table 1). Only standardized achievement tests would remain to be considered seriously. The achievement tests with the best prospects would seem to be those that have the highest *g* loadings, which, if verified in the model, would simply return attention to *g* rather than replace *g*.

If parameters from such a highly *g*-loaded achievement test did provide a good fit, one might have to conclude that general thinking skills rather than mastery of a specific subject was responsible, for the test might turn out to be one of mathematical achievement, for example, or of some other school subject not easily

linked directly to reaching an opinion about the Simpson trial. The thinking skills interpretation would thus again circle back to *g*, and little would have been gained by the attempt to displace *g* in the model. Highly *g*-loaded achievement tests in more than one subject might fit about equally well, such as reading achievement and mathematics achievement—an embarrassment of riches. Measures of information about school subjects that are quite different from one another often correlate quite similarly with intelligence (e.g., Lubinski & Humphreys, 1997, Table 1). To avoid an explanation involving *g*, one would have to explain why both reading and mathematics, and perhaps other topics as well, were unexpectedly about equally relevant to the Simpson case. Whatever the answer, there would be need to invoke a level of abstraction higher than that of the tests themselves (e.g., Jensen, 1984), and so an explanation based on achievement is not a real alternative to *g*. Nor is such an explanation attained by referring to *g* as “achievement,” as is often done (e.g., Neal & Johnson, 1996), or as “basic skills” (e.g., Berlin & Sum, 1988). These substitute terms serve as little more than declarations that authors intend to disregard the substantial heritability of *g* (Plomin & Petrill, 1997).

In contrast to proposals on behalf of unspecified alternative variables, or on behalf of achievement tests that are at once too varied and too specific to connect meaningfully with all the outcomes thus far considered, reasons have been presented in the course of this article linking the various outcomes examined to *g*, the great strength of which is its causal generality and the stability over the relevant time periods of its (relative) parameter values for populations. Besides possessing the requisite parameter values, therefore, a different candidate variable that was concrete rather than speculatively ad hoc would also have to offer theoretical linkages between itself and the outcomes that were at least as plausible as those for *g*, yet different in essence. Until such a candidate materializes, there is good reason to infer that the structure is based on *g*.

#### **UNDERSTANDING WHAT OUTCOMES THE POPULATION-IQ- OUTCOME MODEL MAY FIT—THE DEVIANCE PRINCIPLE APPLIED TO OPINIONS AND VALUES**

The success of the population-IQ–outcome model in fitting certain rumors and opinions does not imply that it can be expected to work on all differences between Blacks and Whites in survey data, just as the more numerous findings in Table 2 do not imply it should work on all behavioral outcomes. For now, indications of what outcomes the model fits are sketchy. A series of successes with prevalences of moderate to severe delinquency and crime having widely varying magnitudes over a long span of time obviously argue for that area. The opinion realm, however, is largely uncharted territory. Judging from the survey items already analyzed, opinions most likely to fit ought to deal with matters of empirical fact

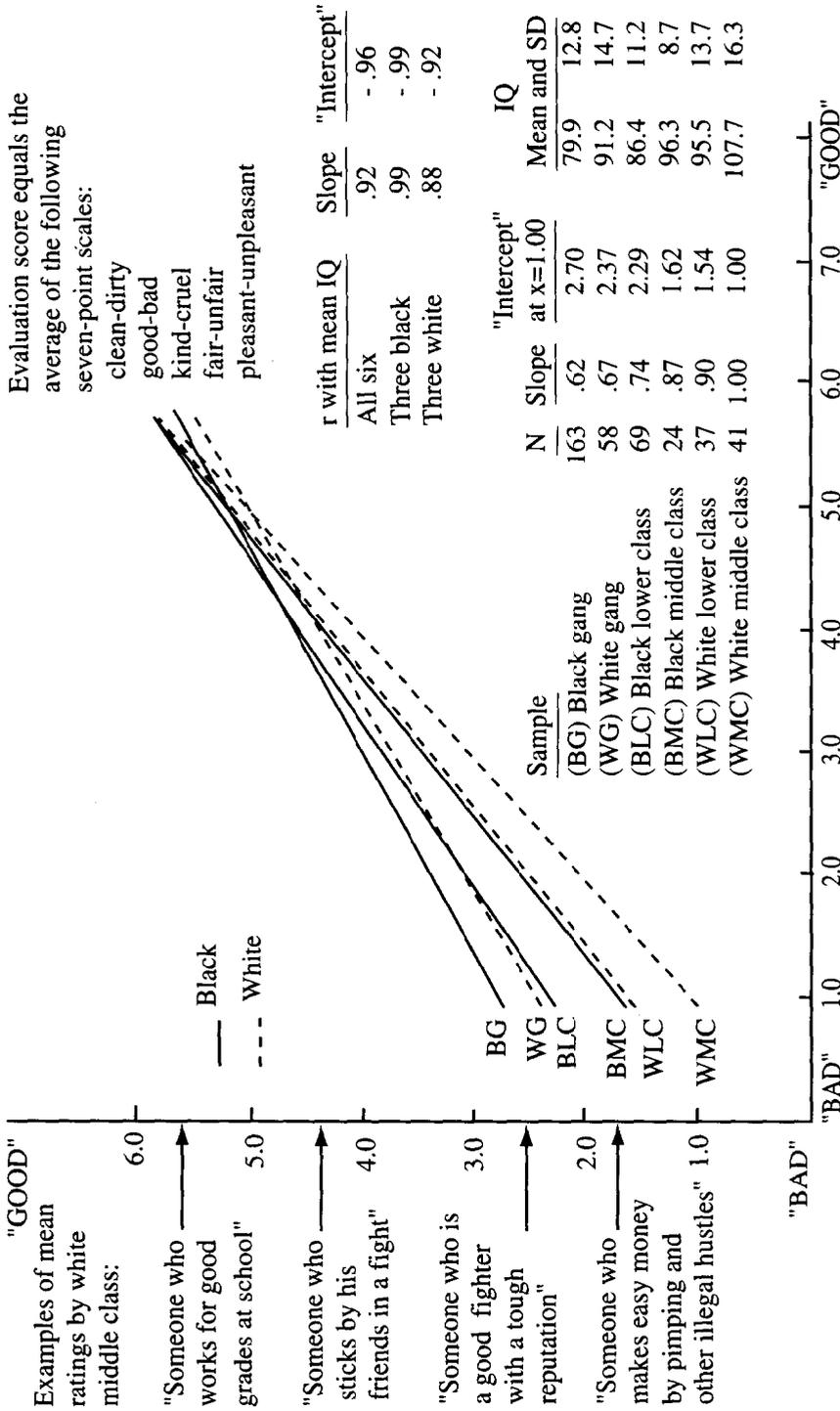
rather than preferences, but this issue remains to be explored further, as even preferences can reflect understandings of fact, as in political preferences. The facts should be either matters of general knowledge (the rumors) or particular matters that are under active public debate and are reasonably understandable by the general public (the Simpson items). "Public opinion takes form in the process of discussion and controversy" (Baur, 1960, p. 210).

The model itself can serve as an exploratory tool for distinguishing among types of opinion, but one that is subject, of course, to the need for validation on fresh material. Examples of poor fit might profitably be considered with the aim of gaining insights into the variables involved, as was the case for SES variables and delinquency (Gordon, 1987) and for the prevalence of nonvirginity among adolescents discussed in connection with Rowe, Rodgers, and Meseck-Bushey's (1989; Rowe & Rodgers, 1994) social contagion models in Appendix C.

In the case of Rowe's prevalence data, it was found that the population-IQ-outcome model fit best for the very youngest adolescents, under age 14. These were the ages at which sexual intercourse might be considered to deviate most from social norms, to be most clearly an "error," if only one of child supervision and, often, of child exploitation. As age increased, and becoming sexually active gradually became less strongly proscribed, the fit worsened. In similar fashion, prevalences for lawbreaking fit well when the criterion did not include many mild offenses that were not so deviant. Like milder forms of criminality (e.g., line 8 in Table 2), single motherhood, which may be frowned upon but which is far less unacceptable than serious crime, displayed consistently negative Black-White differences in critical IQs (lines 13-16 in Table 2), indicating a surplus of Whites from the standpoint of the model. Such a surplus seems to be the hallmark of outcomes that are not highly unacceptable to society.

The various findings reviewed, and additional ones concerning opinion to follow, constitute the basis for *the deviance principle*, the idea that IQ explains group differences better the more deviant (socially erroneous) the behavior in question, holding sample size constant (see also Gottfredson, 1997, Table 10). As outcomes become less deviant, group differences narrow, leaving less opening for intelligence differences to matter. The deviance principle not only accounts for certain important departures from good fit of the model, it also adds to the construct validity of the hypothesis that IQ or *g* is the operative variable through its focus on degrees of error.

Figure 2 extends the deviance principle, which was derived from performance data, to measured opinions by demonstrating that differences in evaluative attitudes (values, opinions) between adolescent populations of varying IQ levels become greater as the subject matter becomes less positively valued and more deviant (socially unacceptable). It also provides another demonstration of the importance of IQ for explaining group differences, in contrast to race and neighborhood SES.



**Figure 2.** Group differences in values and IQ (regression lines, for six groups of boys, of their average ratings of 14 behaviors regressed on ratings of White middle-class boys). From "Crime and Cognition: An Evolutionary Perspective" by R.A. Gordon, 1975, *Proceedings of the II International Symposium on Criminology*. International Center for Biological and Medico-forensic Criminology, Sao Paulo, Brazil. Copyright 1975 by Robert A. Gordon. Reprinted by permission.

Figure 2 was derived from mean ratings of 14 nonjudgmental descriptions of behavior, called *images*, which ranged from highly conforming to highly deviant, by Blacks and Whites who belonged to highly delinquent gangs, lower class non-gang groups from the same neighborhoods as the gangs, or middle-class groups of boys in Chicago (Gordon, Short, Cartwright, & Strodtbeck, 1963, Table 3). The 14 images were thought to epitomize typical subcultural values of each of the six samples, who rated them on five 7-point scales listed in the figure (e.g., clean-dirty, good-bad). Scales were averaged to form scores on an evaluation dimension (Osgood, Suci, & Tannenbaum, 1957; Osgood, Ware, & Morris, 1961). For simplicity, that dimension is labeled "bad-good" in Figure 2.

Four examples of the images (Gordon et al., 1963, Table 2) mark points on the vertical axis corresponding to the average ratings given them by White middle-class boys. The 14 average ratings by each of the groups were plotted against the 14 White middle-class ratings, and best fitting lines were determined (by least-squares regression). Only these lines are displayed, along with their slopes and the points at which they intersect the vertical axis at a rating of 1.0, the lowest possible. Naturally, the White middle class, which serves as the referent, has a slope of 1.0 and an intercept of 1.0 when regressed on itself. The choice of group to serve as referent is a matter of interpretative convenience and does not affect the results of interest here.

Differences in subcultural values represented by differences among the lines in Figure 2 are thus quantified in two different ways, one using slopes of the lines, the other their intercepts. The lower the slope or the higher the intercept, the less inclined a group was to condemn deviant behaviors by rating them in varying degrees as "dirty, bad, cruel, unfair, and unpleasant." The well-ordered nature of the ratings summarized by the lines indicates that, on average, a group that is more tolerant of one form of socially unacceptable deviance will also be more tolerant of other forms.

Mean scores on a nonverbal test of IQ (Cattell & Cattell, 1958) were available for each group, and these also appear in Figure 2. The IQ means of Black (79.9) and White (91.2) delinquent gang members in the early 1960s remain typical of those of delinquent Blacks (79.1) and Whites (92.8) committed to a training school for males two decades later (Simonds & Kashani, 1979, p. 1446). Although IQs of 79 may seem too rare to matter much to society, 79 was in fact the mean IQ of children from the lowest SES stratum, constituting a clear majority of the entire normative sample, in the classic study of Black elementary school pupils in the Southeastern United States by Kennedy, Van De Riet, and White (1963, Table 42).

The important observations concerning Figure 2 are these. In the upper right corner, all six lines converge at the highest mean ratings actually observed, between 5 and 6 on the scale. There was thus general endorsement, by all six sam-

ples, of middle-class behavioral images that fell in the uppermost "good" part of the attitude range, and no significant differences among samples were found. In the lower left corner, however, the "bad" part of the range where the most delinquent images of all fell (mean ratings of pimping ranged from 1.8 to 3.5), the lines splay farthest apart, and significant differences were often found between the groups' ratings of single images (Gordon et al., 1963, Table 4). Behaviors of intermediate goodness and badness fell along the lines between the two corners mentioned. Each other group's line seems to swing upward and away from the line representing the White middle class, who rated the deviant images lowest of all the groups. The lines for gang boys swing farthest of all.

Whether or not race is held constant, the correlations between mean IQs, on the one hand, and slopes and intercepts, on the other, are so high that it could be said that IQ accounts for almost all of the variance in the two statistics used for summarizing the linear profiles of subcultural values. Average intelligence accounted for 85% and 92%, respectively, of variance in the six groups' slopes and intercepts for values. In two cases in which pairs of groups have relatively similar IQ means, less than five or less than one point apart, they have similar intercepts even though not of the same race. IQ, not race per se, appears to be the operative variable in producing group differences in these subcultural values, and those differences are largest for opinions about the most deviant descriptions of all. Within race, differences in IQ of about four to six points are associated with differences in values between gang and nongang boys from the same neighborhood settings.

The upward swing of groups away from the line of the White middle class in Figure 2 can be related to what the sociologically minded Senator Daniel Patrick Moynihan (1993, p. 17) called, "defining deviancy down." Moynihan's well-known phrase meant that deviant acts were being viewed more and more tolerantly in American society; thus, his down becomes up in Figure 2. In terms of Figure 2, Moynihan's argument is comparable to claiming that all social groups, in varying degree, are adopting, or at least tolerating in others, lines that swing up toward that of the group that is most deviant, that is, the group that departs most in both behavior and attitudes from what has long been considered socially acceptable. The strong correlations between the locations of the lines and IQ indicate, however, that it is accommodations to group differences in  $g$  and to the effects of those differences that may be driving the seeming redefinitions of deviance that Moynihan had in mind.

That delinquent gang members do not differ from the White middle class in their evaluation of middle-class images at the upper right corner of Figure 2 may seem puzzling. After all, should not deviants deviate? The problem arises from the fact that deviance and conformity are often mistakenly viewed as mutually exclusive, and so distinctions become muddled. If the part of the deviance-con-

formity continuum on which behaviors are located is kept distinct, however, as in Figure 2, it becomes evident that deviance and conformity are largely not mutually exclusive, and they are only so at any one point on the continuum. One can both deviate and conform, but in different ways, and so get intellectually easier opinion items right and more difficult opinion items wrong, just as with items on psychometric tests. Even the worst criminals conform most of the time (e.g., wear clothes outdoors). The additional ingredient of temptation in deviant choices often presents individuals with a harder intelligence item than mere conformity without temptation to deviate. For painless conformity, the answers are often so easy to figure out that hardly anyone chooses wrong.

Another muddle arises from the common practice in social science of viewing achievement as conformity (to middle-class norms). Lack of achievement thus becomes nonconformity automatically, and so it may seem inconsistent that gang boys, who would be low in average school achievement (and thus nonconformist), espouse conformist attitudes in favor of achievement. The more one relies on motives and culture to explain all behavior, the more troubling is this seeming inconsistency. Individuals can, however, share high achievement norms (and thus conform) without being able to achieve highly, often for IQ-related reasons. It would seem inappropriate to label nonachievement for such reasons *deviant* in the same sense as more voluntary forms of socially unacceptable behavior. Errors in the case of achievement consist more often in not being able to perform the right behaviors, at least not easily enough, rather than in not choosing to perform the right behaviors. Much of what passes for social science and public policy depends on blurring this fundamental distinction.

Admittedly, the deviance principle as a guide to what is grist for the population-IQ-outcome model leaves much ground uncovered, especially in the opinion realm, but it represents a strong heuristic, as Figure 2 shows, with demonstrated applicability to at least three different forms of outcomes considered so far (age of first intercourse, lawbreaking, and, beyond the model, values). Consistent with the principle is the fact that there is often little difference in opinion between Blacks and Whites on core values, such as respect for tradition and for religion (e.g., Ladd, 1996, pp. 38–39) or on what Gordon et al. (1963) called *prescriptive norms* as opposed to the *proscriptive norms* meant to restrain misbehavior. Not everyone can succeed equally in educational achievement, but everyone can understand that education is a good thing (“. . . works for good grades at school”). There is also no race difference on certain metaphysical topics, beyond reach of empirical verification. About equally large majorities of Whites and non-Whites, for example, say they believe in heaven, hell, and the devil (McAney, 1995). The social unacceptability of an opinion, and not just its lack of factual basis (which may diminish acceptability under debate conditions), appears to play a major role in determining fit.

## DISCUSSION AND CONCLUSION

### **The Neglect of General Intelligence**

In the social sciences, noneducational behavioral outcomes, such as criminal status, single motherhood, HIV infection status, and opinion status are often subjects of systematic inquiry. Theories based on motives, values, social learning, culture, social structure, economics (poverty), and power, with rare exceptions, have effectively dominated attempts to explain such outcomes.

Investigations taking account of intelligence are relatively infrequent, and they have succeeded best in gaining a foothold where intelligence test data for individuals were available to be joined with individual outcomes in within-group models, as happened in the cases of delinquency (Hirschi & Hindelang, 1977) and job performance (Gottfredson, 1986b; Hunter & Schmidt, 1982). Attempts to add *g* to explanations of group differences (Gordon, 1976; Gottfredson, 1986a) have aroused more resistance. Herrnstein and Murray's (1994) *The Bell Curve* exploited an unusual data set that happened to include scores on a good test of *g* with records on a variety of individual outcomes, but reactions to their work have often been dismissive (Fraser, 1995; Jacoby & Glauberman, 1995; Kincheloe, Steinberg, & Gresson, 1996), as though their findings were merely empirical or incidental rather than possibly causal associations, and overstated at that if not the products of misanalysis (Gould, 1994; see Gordon, 1995). Ironically, Herrnstein and Murray's basic model is a within-group one, and thus typical of much sociology except for the respectful treatment given *g*. Hence, their measures of effect size often fail to convey the greater importance that *g* can assume at the population level.

### **Illustrating the Underrecognized Importance of Intelligence**

This article has attempted to address the seemingly entrenched skepticism behind negative reactions to explanations based on IQ. First, the article began by drawing on elementary principles governing intelligence test items, with the view that many behavioral outcomes can usefully be understood as analogs of single test items in psychometrics. Popular confusions about the role of intelligence in those outcomes would be subject to clarifications similar to those that account for poor results under similar conditions from regular test items considered singly. The aim was to reduce skepticism about the possibility of an important role for intelligence in everyday behaviors that often colors the reception even of good data such as Herrnstein and Murray's, and allows criticism, no matter how ill-conceived and uninformed, of research like theirs to seem persuasive to other intellectuals and perhaps to the general public as well.

In a series of three steps, the article next laid an empirical and theoretical basis for understanding better the role of intelligence in explaining differences in major

behavioral outcomes from everyday life. Absent a deeper understanding of how intelligence works, its effects remain underrecognized in this realm. The three steps started with the level of individual probands, moved to the social context of the probands, and culminated with differences between entire racial populations using a population–IQ–outcome model that unified all three levels by pooling their effects. The model was tested on two well-defined populations with different IQ distributions using data that spanned 50 years.

Illustrations were provided at the individual level of how intelligence can usefully be viewed as the probability of (not) making cognitive errors. The emphasis on individual differences in probability, even for outcomes in everyday life not readily explained by schooling (e.g., the Simpson opinions), brought out key issues that are often not well understood: Intelligence is more fundamental than just the knowledge one has accumulated by whatever means; the probability of error depends jointly on the difficulty of the task involved and one's IQ level (cf., Gottfredson, 1997); probabilities of error for a given IQ need not be zero or 1, but only intermediate (like batting averages), to be important in a practical sense.

The second level of analysis, the contextual, illustrated how probabilities of error (or success) are further modulated by the intelligence level of one's near context, particularly family and peers. Explanations of differences in *g*-related individual behavior must thus take account not only of individuals' own levels of *g*, but also the levels of the individuals who form their proximal environment. Functional advantages go not only to individuals who are brighter, but also to those who are fortunate enough to live among brighter individuals and, especially, to participate in the *fraternité* of acceptable reciprocal exchange with those individuals. This second level of analysis also shows that group processes, such as the diffusion of dependable information and sound opinions, may often be better understood by attending to the intelligence context formed by the social groups in which the process occurs.

Variables measuring immediate social environments often appear in explanatory models as sociological ones (e.g., differential association with delinquents, parental SES), but their systematic contributions to probands' outcomes could be the result of the mutual relationship between their own complex relations with intelligence and the intelligence of the proband (itself too often left unmeasured). In the new perspective offered here, for example, scholastic underachievement in relation to probands' IQs that results from disruptive behavior by classmates is not just a sociological effect, but quite conceivably one related to the IQs of the classmates that adds to the sum of effects of *g* on scholastic performance. HIV infection is a reflection not only of one's own judgment, but of the judgment of one's sex and needle partners, and of their partners.

If *g* is not to be sold short, care must be taken, when considering its effects, to qualify certain forms of intelligence data as applying only to the individual *g* of probands (as distinct from data reflecting, in addition, the local intelligence con-

text, or, ultimately, the potential full effects of  $g$  in the probands' population on prevalence of the outcome concerned). For example, the phrase supplied in brackets needs to be added to statements such as the following, which summed up the fact that Herrnstein and Murray (1994) had reported unexplained residual differences between races in certain undesirable outcomes after controlling for individual IQ: "Given that IQ was equated for Blacks and Whites, it is clear that much more is contributing to differences in societal outcomes than just IQ [*of the proband*]" (Sternberg, 1996, p. 15). Most of the residual differences that prompted this critical comment, when portrayed as percentage differences instead of ratios of two, three, or five to one, were in fact rather small, 8% or less. The largest percentage difference (and source of the 5:1 ratio), 41% for out-of-wedlock births, obviously involved partners, not to mention wider contexts, whose IQs were *not* controlled (e.g., Dearden, Hale, & Woolley, 1995).

The recognition of a relation between intelligence and intermediate group-level phenomena brought us to the third level of analysis, the population level. Two well-studied populations, U.S. Blacks and Whites, were used to demonstrate population-level effects of differences in intelligence distributions. A population-IQ-outcome model was described that incorporates the effects of  $g$  at both the individual and contextual levels, including now remote contexts within the population, such as spokespersons, elites, leaders, and chance encounters. The model was then applied to a wide variety of outcomes for Blacks and Whites in order to test whether population differences in IQ distributions can explain differences in prevalences of various social outcomes. Outcomes concerning juvenile delinquency, adult criminality, single motherhood, HIV infection, poverty, conspiracy rumors, and two forms of key opinion concerning the O.J. Simpson case were found to be commensurate with differences between Blacks and Whites in IQ distributions. The model thus accounts for the presence of good as well as bad outcomes in both races, by addressing successfully the important question of the *difference* in their relative levels within each race, a question that race per se has never been able to answer. Racial categorization itself, although useful for applying the model, is thus seen to have rough descriptive rather than explanatory value, as IQ appears to be a more fundamental variable. Subcultural values and attitudinal toleration of socially unacceptable deviance were also found to be related to average IQ.

The striking success of the population-IQ-outcome model for such a variety of outcomes argues that intelligence is an active if much less exclusive cause of the same outcomes at the individual and contextual levels too, even though corresponding IQ data in most cases were unavailable for those levels. IQ or  $g$  appears to function as a master variable exerting important effects on differences between populations on a number of different kinds of variables, just as socioeconomic status and culture have long served as the putative master variables of sociology and anthropology, respectively. Numerous small effects, often invisible to the

unaided eye, of intelligence at lower levels seem to eventuate in large effects at the highest level, a situation not unknown in other branches of science.

Analyses demonstrated, even in the absence of individual-level data on IQ, that the model can assess the possible importance of  $g$  in explaining population-level differences in outcomes of behavior. Lack of large data sets containing IQ measurements has been a barrier to progress in exploring the role of intelligence in social life, but the model allows progress to be made in their absence, as was demonstrated, for example, with the Simpson and Brawley surveys. Both successes and failures of the model can lead to new hypotheses for explaining the patterning of population differences in outcomes. For example, when, within a particular domain, the model fits different opinions about equally well despite their having different prevalences of occurrence, as in the case of the unequally endorsed rumor items in Table 3, one can surmise that those opinions involved problems at different levels of difficulty. More obviously wrong opinions, for example, will be held by (got wrong by) a smaller percentage of persons in both populations. Thus, an intellectual difficulty hypothesis has been extended to the opinion domain by certain successful tests of the model.

Another, closely related, example of new hypotheses was the explanatory deviance principle developed from a pattern of certain failures of the model. That principle is that the model fits more deviant or more socially unacceptable acts within a domain that lends itself to such distinctions better than it fits less deviant or nondeviant acts, that is, bad mistakes tend to fit better. The principle is most efficient when used to exclude outcomes as candidates for the model.

The use of different racial distributions served as a device for exploring effects of intelligence at the level of populations, and revealed that it was not race per se but  $g$  that could account for many important differences typically associated with racial classifications. Following a certain amount of expectable early success, in view of outright discrimination in the past, the long-standing focus on race is proving increasingly unproductive for understanding what are still conceived of as "racial" problems. As frustration mounts, that focus, unfortunately, is becoming increasingly potent as a polemical device.

### **Implications Concerning the Central Role of Intelligence**

The success of the population–IQ–outcome model in fitting a considerable variety of outcomes warrants taking it more seriously as a description of reality than would be the case if it remained just a method for detecting a likely IQ effect on delinquency. Certainly, the model represents a radically different way of looking at social phenomena, but one that is consistent with the human ecology school that assigns importance to population "as an aggregate of biological units" (Hawley, 1950, p. 77). The question, How does a society work? is partly replaced by the new question, How does a population work? The answer is still: In very complicated ways. But now some, at least, of those ways seem to be coordinated

or orchestrated to a surprising extent by the population's distribution of  $g$ . The variable  $g$  is a widely influential and relatively powerful one that simplifies the calculations and reduces the number of key inputs needed to discover regularities between it and the prevalence of many social outcomes. In fact, it may be the pervasiveness of the effects of  $g$  more than their absolute power when looked at individually that does the trick. According to the various forms of evidence, there would appear to operate a law of (combined) averages for the various effects of IQ, which leads those effects to aggregate in large populations and produce the surprising regularities between prevalences of many outcomes that have been reported here.

Such a dependable form of aggregation is not as bizarre as it may sound, because there have been limited hints that such regularity is possible when sample size is large. It would not be considered surprising, for example, if differences between many populations on a 60-item test of scholastic achievement were highly predictable from their differences on a 50-item IQ test, even if the two tests were administered to separate representative samples of those populations at somewhat different times. Yet, if one thinks about it, a huge number of minor details would have to go just right, on average, and on a regular basis, within each population for such a result to appear.

Percentages (proportions) are, after all, just another kind of average, obtained when the two possible outcomes have been scored zero or 1. Because the measurement properties of the behaviors behind the population percentages are not as favorable to  $g$  as the measurement properties of the behaviors behind population averages on psychometric instruments, much larger samples might be required for observing the same high degree of regularity in percentages that is more or less expected of average scores on  $g$ -loaded tests.

The ubiquitous consistency of  $g$  is hinted at by correlation matrices and regressions involving ability and sociological measures that often appear to differ little across race when analyzed comparatively (Gordon & Rudert, 1979; Rowe, Vazsonyi, & Flannery, 1994). Race differences in key statistical minutiae of criminal activity, such as frequency of offending for offenders and arrests per year per active offender, also appear negligible (see Appendix C). The familiar observation that the correlations among variables (especially among outcomes such as education or crime) in social science tend to be almost entirely positive in sign (or can be made so by suitable reflections of arbitrary scoring directions) may be simply a weaker case of the classic positive manifold (positive intercorrelations) shown by cognitive tests when factor analyzed (e.g., Maller, 1933, pp. 116, 121). Sociological and cognitive correlation matrices may be alike in that they reflect, but to different degrees, the presence of  $g$ , the general intelligence factor found to be the same  $g$  factor for Blacks and Whites (Jensen, 1985). Presence of the  $g$  factor is especially apparent when variables in a sociological correlation matrix happen to be aggregated ones, such as the education, income, and prestige values associated

with specific jobs (e.g., Gordon, 1987, pp. 49, 86–88; see also Gottfredson, 1985, p. 141, 1997, Table 7). Occupational prestige, recall, was found to be highly correlated with the intelligence demands of jobs by Duncan et al. (1972, p. 77), and loadings on a prestige factor can account entirely for the correlation between education and income attributes of jobs. Much of social structure appears to be organized along various axes of human instrumental activity (e.g., mental, athletic, economic); to the extent differences in instrumental performance cannot be eliminated or glossed over, social structure as we know it will tend to persist.

The difference between the population–IQ–outcome model and the usual sociological approach to explaining race differences can be likened to two different approaches to explaining the cracks that radiate from a single point of impact to a mirror. The traditional sociological approach notices the cracks and attempts to use some of them to explain the others. Explanation often begins at any point, and can conceivably go round full circle in either direction if followed through the hands of different theorists: “one kind of pathology breeds another” (Clark, 1965, p. 81). Pessimistic attitudes of teachers, it is alleged, cause low achievement, rather than vice versa (Clark, 1965, pp. 128–129). Poverty causes low IQ, rather than vice versa (Hurley, 1969). Poverty causes crime (Bourgois, 1995), and crime causes poverty (J. Lewis, 1995; Rubinstein & Horowitz, 1995). Closeness of cracks, and their convergence toward the point of impact may lend special cogency to some attempted explanations. In tacit acknowledgment of confused efforts, sociological explanations have sometimes employed phrases such as “the tangle of pathology” (Moynihan, 1965, p. 29) and “chronic, self-perpetuating pathology” (Clark, 1965, p. 81) to convey the lack of any clear causal progression, for “the roots of the multiple pathology . . . are not easy to isolate” (Clark, 1965, p. 106). Inevitably, metaphors such as “poverty cycle” and “vicious circle” are invoked (Zarefsky, 1986, pp. 106–107). Not surprisingly, social scientists find themselves trying to counter the impressions that “nothing works” (e.g., Tonry, 1995, p. 201) and that an “entire culture seemed impervious to modification” (Zarefsky, 1986, p. 107).

The population–IQ–outcome model finds that population IQ differences can represent the single impact and thus explain many of the cracks heading away from it. Some historical explanations have represented efforts to sidestep the circle dance by identifying a single impact event, such as the period of slavery, that would account for the many cracks (e.g., Fleming, 1976; Moynihan, 1965). An important difference between the legacy of IQ and the legacy of slavery, however, is that the former can be quantified and used to explain quantified outcomes, the latter cannot.

### **Implications Concerning a Deeper Structure**

Evidence that differences in  $g$  can have pervasive large effects between groups is important, but more important perhaps is that such differences can, remarkably,

account for a complex pattern of results with deep implications. Thus, the good fit of the population–IQ–outcome model to a great variety and number of outcomes, ranging in level of response (prevalence) from almost zero to over 95%, at widely separated points in time, delineates and establishes the existence of a stable, deeper structure than is ordinarily visible in social data, whatever its nature. Even Bailey's criticism of the model, the only serious attempt advanced thus far, acknowledged as much. Some 44 or so instances of good fit to IQ, not all independent, have been presented in this article (excluding some as too interdependent). The likely nature of the structure they imply was indicated in the excellent result from the Simpson surveys, where it was inferred that the means and variances of a normally distributed variable such as  $g$  defined the underlying structure for the two populations in question. Prevalences for Blacks and Whites of fairly miscellaneous other outcomes, delinquency, adult crime, single motherhood, HIV status, poverty, and conspiracy rumors, also appeared to operate in tandem with the parameters of the two IQ distributions, and so bolstered the IQ inference concerning the Simpson data by reaching beyond the fact of good fit in any one kind of outcome. It was as though those prevalences were tracing the outlines of overlapping but separated normal distributions of a deep causal variable, subject to the resolution limits of the data and occasional disturbances. Averaging results from many applications is thus akin to superimposing many such tracings in order to obtain a composite picture.

Gottfredson's (1986a) model of expected Black–White differences in employment in nine major occupations, based on Black–White differences in IQ distributions, is consistent with the model presented here. Real occupational data do not lend themselves to confirming the model as readily as the variables considered in this article, however, because hiring can be subject to disturbances, such as affirmative action and, in the past, net discrimination against Blacks. Such disturbances were reflected in Gottfredson's results, but in a manner that did not interfere with a key conclusion, namely, that the relative degree of disparate impact in employment across occupational levels can be predicted by the  $g$  demands of the jobs and the  $g$  distributions of Blacks and Whites. Other outcomes that the population–IQ–outcome model fails to fit precisely might profitably be investigated for weaknesses in data or for new insights, rather than dismissed as being unaffected by IQ. Data series that manifest trends linked by a change in fit can profitably be examined with an eye to relevant changes in historical circumstances.

### **Implications Concerning Alternative Forms of Intelligence**

It might reasonably be expected that Black–White differences in forms of mental ability theorized to be independent of  $g$  or independent of so-called academic intelligence (Ceci & Liker, 1986, p. 119), such as practical intelligence (Sternberg & Wagner, 1986) and multiple intelligences (Gardner, 1985), would enter into the

explanation of Black–White differences in various “run-of-the-mill” (Goodnow, 1986, p. 86) behaviors that have such practical consequences as crime, HIV infection, poverty, and opinion on controversies of the day. According to Sternberg (1987, p. 245), for example, “The ultimate test of intelligence is one’s ability to apply one’s mental skills in one’s everyday life.” “Stated simply, [intelligence] is mental self-management” (Sternberg, 1988, p. 72). Depending on which race held the advantage in non-*g* abilities related to “competence in the everyday world” (Sternberg & Wagner, 1986, p. iii), perhaps because one race offered a more “supportive cultural context” (Gardner, 1985, p. 317) than the other for the development of such abilities or caused them to develop in ways that were unrelated to *g* (Ceci & Liker, 1986, p. 138), the Black–White gap in outcomes would be reduced or enlarged beyond what *g* might explain.

Everyday intelligence has also been thought to be more context specific than academic intelligence (Biggs & Collis, 1991), and the outcomes named above are certainly diverse enough to permit specific forms of intelligence to flourish in connection with each. Widely studied moral reasoning ability, for example, might be thought likely to affect delinquency and crime, if not the other outcomes, independently of *g*. Recent findings are, however, that even at the individual level moral reasoning adds no incremental validity to *g*-loaded verbal ability for predicting 62 various outcomes (Sanders, Lubinski, & Benbow, 1995). We can name more things than we can measure independently.

The consistently good or excellent fit of the model based solely on *g* obviates the need for positing additional explanations of differences *at the level of populations* for the various outcomes and two populations examined in this article, with the possible exception of single motherhood, which had a small but consistently directional residual. It is reasonable to expect that the roles, if any, of relevant non-*g* intelligences and abilities at the individual level would have been systematically magnified at the population level so as to spoil the fit if such variables were present to different degrees in each race. The possibility of non-*g* abilities being unequally present for Blacks and Whites remains a key contextual assumption of research on practical intelligence: “Intelligence is not quite the same thing for any two groups” (Sternberg, 1987, p. 246). Social competence (Ford, 1986, Table 2) or interpersonal intelligence, the latter in particular thought by Gardner (1985, p. 317) to be relevant for addressing juries, should find their natural milieu at the level of social contexts and thus influence population outcomes at the next level up. But this did not appear to happen either, as *g* left no residuals to speak of, plus or minus, in need of additional explanations that could not reasonably be dismissed as noise (but for single motherhood and, in later years, poverty). Either Blacks and Whites are nearly equal in the non-*g* abilities relevant to the outcomes examined in this article, which would largely confine their possible effects to within-group models, or different mixes of abilities in each race have offsetting values and hence no additional net effect between races on the variety of outcomes

considered (an unparsimonious assumption), or the non- $g$  abilities in question do not exist independently of  $g$  in the general population to a detectable degree.

The logical possibility remains open that non- $g$  abilities do figure in other outcomes or in other populations that the model will not fit well, although their elusiveness in the populations and outcomes considered here is hardly encouraging. The model can be used as a point of departure for detection work aimed at discovering why it fits well in some cases but not in others, with the expectation of learning more about the latter cases by asking new kinds of questions, as were raised for SES variables and answered by empirical tests of the IQ-surrogate hypothesis (Gordon, 1987). Conceivably, such detection work could uncover non- $g$  sources of mental ability in between-groups effects, although it is more often likely, I think, to lead in the direction of explanations rooted in social (equal pay for workers in the same job), legal (affirmative action, minimum wage laws), cultural (religious attitudes concerning sex, in-group slang), and communicative (imperfect diffusion, indifference to many issues) phenomena, that can lead to racial and ethnic differences in outcomes that are too small for the model to fit, but sometimes, conceivably, also too large. Such disturbances to the model need not be the same for every pairing of racial or ethnic groups.

### **Implications Concerning Higher Order Consequences of Intelligence**

The importance of a variable such as  $g$  resides not only in its effect size, whether absolute or relative to that of competing explanations, at any of the three levels, but also in its practical consequences, that is, the kinds of effects in question. Many of the behavioral outcomes examined, such as crime and HIV status, are obviously consequences of great importance in their own right. These consequences themselves have second-order consequences, however, as they each become part of the context for one another's outcomes.

Efforts have been made to document and spell out second-order consequences for the opinion outcomes concerning both rumors and the microcosmic Simpson trial, because the novelty of those variables might render their further consequences less apparent than those of crime and HIV, which, at the very least, produce second-order consequences through familiar first-order economic costs. Like all social problems, crime and HIV infection obviously compete with each other to some degree for resources needed to combat them both, and thus they potentially decrease the effectiveness of society's response to either or both of them, to name just one set of second-order consequences for those outcomes. The average cost of state imprisonment in 1992 was about \$19,500 per inmate-year (DiMascio, 1995, p. 9), for example, and the cost of AIDS care in North America in 1990 was about \$32,000 per person-year (Tarantola, Mann, Mantel, & Cameron, 1994, Table 16). In 1990, the total cost of corrections reached \$24.9 billion (DiMascio, 1995, p. 8), and the combined costs of care and prevention of AIDS came to \$2.5 billion (Tarantola et al., 1994, Table 16). Calls to reduce expendi-

tures for imprisonment (e.g., DiMascio, 1995; Mauer & Huling, 1995) and to increase expenditures for HIV control (e.g., Lambert, 1988) are frequently heard. These are but two of the many possibilities. The impaired immune systems of HIV-infected persons have also potentiated tuberculosis epidemics in many parts of the world, reversing gains made against that disease over the years (Leary, 1995).

The rumors, recall, had dysfunctional repercussions within the Black community. The Simpson verdict, which may be understood best, perhaps, as a special case within the wider context of public opinion—and of public problem solving revealed by polls—is producing consequences not all of which are yet evident. Some could entail significant institutional changes in the jury system, for example, which itself might have unforeseen consequences on attitudes toward the justice system (e.g., Bugliosi, 1996, pp. 286–287). Johnnie Cochran has already commented, “Now that we understand the rules, some people want to change them” (Bruni, 1996, p. B3). Of the population at large, 38% felt that the jury system is in need of major change following the verdict (Moore & Saad, 1995). The verdict itself seems to have had no immediate deleterious effect on racial animosity as measured by recent trends in opinion polls (Moore & Saad, 1995), despite widespread convictions in both the Black and White communities that both it and the accompanying poll data would hurt race relations (Bugliosi, 1996, pp. 278–280; Darden, 1996; Milloy, 1995; Moore & Saad, 1995; Rowan, 1995; Shapiro, 1996, p. 355). In that extreme respect, at least, Whites seem to be taking the verdict in stride, notwithstanding dire predictions (e.g., Milloy, 1995). This result might be kept in mind whenever such predictions are used to justify prior censorship of research findings.

Moynihan’s famous *O tempora, o mores* can be viewed as a response to second-order effects unknowingly driven in many cases by first-order effects of intelligence differences. What he called defining deviancy down can be construed as just one of many sorts of higher order accommodations in culture, law, and social policy that are brought about by *g*-related group differences in behavior. When differences in IQ distributions are stable over time, as the Black–White difference has been (Gordon, 1980b, 1986, Table 2), there will be predictable race differences in prevalences of *g*-influenced behavior unless those effects are somehow blunted (Herrnstein & Murray, 1994, chap. 14). Blunting effects has proved difficult in the respects with which I am familiar (consider adolescent pregnancies, drug use, Black-on-Black homicide, school achievement, child abuse; all of these have been targets of intensive campaigns). Only poverty has reflected a limited success in reducing IQ commensurability, and that mainly because money and jobs can, to a certain extent, be redistributed. Tables 1 through 5 and Figure 2 showed, in fact, that Black–White differences in diverse outcomes could often be accounted for entirely (delinquency, crime, HIV infection, poverty, opinions) or almost entirely (single motherhood, values) in terms of differences in *g* distributions. Not only were these race differences predictable, therefore, they were often totally predicted by *g* distributions.

When policymakers attribute such differences in prevalences to properties of the larger society without regard to differences in the properties of the populations themselves, there occurs a shift in emphasis from errors made by members of the population to errors made by the society or system that in itself constitutes a redefinition of deviance. Sociological labeling theories, which are more concerned with who defines certain outcomes as deviant than with what causes the behavior so defined, are a prime example of the shift in emphasis (e.g., Mercer, 1973, on which see Gordon, 1975/1980a, 1980c, 1984, Gordon & Rudert, 1979; Quinney, 1970, on which see Gove, 1980, and Tittle, 1975/1980).

It is no coincidence that it was a labeling theorist who first proposed what amounted to the race norming of mental ability tests (Mercer, 1973, 1979), a redefining of departures from the norm for one group (when diagnosing mental retardation) that later reappeared in the employment testing realm as a device for redefining occupational competence (Gottfredson, 1994). Race norming of job tests amounted to a double standard that was recognized to be far less costly in near-term lost productivity than a single, but lower standard of job qualifications that would apply to all persons hired (Hunter et al., 1984, p. 94). Such a lowered single standard would have represented the complete analog to Moynihan's observation concerning redefinitions of deviancy. Partial or complete, the parallel is there. The societal choice when trying to eliminate adverse minority impact in such IQ-driven outcomes seems to be between a universal but costly lowering of standards for everyone that can amount to the dismantling of a society's culture (e.g., Gordon, 1988), and an uncomfortable, perhaps ultimately untenable, but less costly double standard. Neither augurs well.

Worse yet, Moynihan (1993, pp. 25–26) described a more insidious development: "A large increase in what once was seen as deviancy has provided opportunity to a wide spectrum of interest groups that benefit from re-defining the problem as essentially normal and doing little to reduce it" (see also Rainwater & Yancey, 1967, p. xi). Thus, an initial difference in deviance between two groups may sometimes serve as the entering wedge of permissiveness in formal or informal policy that leads to an increase in deviance in both groups over time, although rising prevalences in both may maintain their commensurate relation to the difference in *g* (e.g., single motherhood).

In effect, deviant behavior, if correlated with race by virtue of IQ differences, can thus expand the definition of what is socially acceptable or tolerable among elites, government officials, and eventually even the resigned general public, regardless of whether or not any of these groups join in the behavior. When a Black holdup man shot a frightened cashier three times before her horrified shoppers at an upscale Baltimore supermarket recently, a resigned manager, surely mindful of that heavily Black city's crime situation, said, "This is the way things are nowadays" (Hermann & Myers, 1996, p. 2B). Indeed they are, and so was his understandable reaction. Absence of once-appropriate strong reaction has been noted

whether victims are Black or White (e.g., King, 1996). Focus groups conducted for the prosecution in the Simpson trial revealed that Black women were far less outraged than might be expected by wife beating, and a telephone poll revealed that "a full 40 percent of black women felt that the use of physical force was appropriate in a marriage" (Toobin, 1996, p. 193). Apparently, they had, as victims, and quite conceivably as perpetrators, long since defined this deviancy down, believing "every relationship has these kinds of problems" (Bugliosi, 1996, p. 94).

The tacit expansion of what is acceptable or of what must be endured can occur even in the absence of assigning responsibility to society, simply as a matter of pragmatic adjustment and avoidance of touchy issues. A suspect in the supermarket holdup and his companion were quickly caught, but such limited reactions fail to address the global condition the store manager recognized all too well. In the society at large, it is becoming more and more apparent that no one is minding the store. Pundits are beginning to remind us that the first responsibility of government is to maintain public order (Magnet, 1993; Thomas, 1994; Walinsky, 1995), a condition that had long been taken quite for granted except during episodes of civil disorder.

Much of Moynihan's concern was directed at what he considered relatively passive responses by politicians to increases in multiple forms of deviance so great that they and other elites seemed to prefer defining away as much of the problem as possible, so to speak, to facing the daunting challenge of trying to control the epidemic effectively. Moynihan made much of what he and others viewed as a weakening of outrage, or at least of public outrage. He did not link that loss of expression to the likelihood of being labeled racist for directing attention to any social problem that is differentially present in the Black community, although he no doubt understood that well from his own experience following his attribution of special problems faced by Blacks to the disproportionate decline of intact Black families (Edsall & Edsall, 1992, pp. 54–55; Rainwater & Yancey, 1967, p. 262; for other such examples, see Taylor, 1992, pp. 80–81).

Moynihan's lesson on redefining deviance teaches anew that all second-order consequences of the various outcomes flow eventually toward the same sea. There they join, producing third-order consequences in the body politic as citizens in both races register the impacts of multiple second-order consequences and react to them with whatever understanding they can bring to bear and with whatever information they have been provided. This is clearly the realm of racial politics (e.g., Edsall & Edsall, 1992; Hacker, 1992, chap. 12; Kinder & Sanders, 1996), but some, at least, of its deeper wellsprings may now stand more clearly revealed. Viewing intelligence and its lower-order consequences as key determinants of racial politics contrasts most sharply with approaches that attribute Black–White differences in political opinions to White racial resentment and that dismiss beliefs in a Black–White intelligence difference as an unfounded racial stereotype, as though the concerns of Whites had no legitimate basis (e.g., Kinder & Sanders, 1996). Disciplinary divisions of labor that lead to investigating race differences in

outcomes singly, because they fall within the provinces of separate specializations, usually ignore implications for third-order consequences (Gordon, Lewis, & Quigley, 1988). The third-order consequences of stubborn group differences in *g*, whatever they may turn out to be, could prove to be the ultimate demonstration of the underestimated importance of intelligence in everyday life.

**Acknowledgment:** Linda S. Gottfredson, guest editor of this special issue, provided especially relevant references and insightful comments, as well as extremely helpful editing suggestions. Her valuable contributions are acknowledged with gratitude.

## REFERENCES

- A guide to the jury. (1995, October 9). *Newsweek*, pp. 32–33.
- Achenbach, J. (1995, October 5). Lawyers' sniping destroys illusion of defense team unity. *Washington Post*, pp. A25, A27.
- Adams, L. (1995, October 5). Simpson jurors cite weak case, not "race card." *Washington Post*, pp. A1, A26.
- After the grand jury. (1988, October 9). *New York Times*, pp. 8E–9E.
- Akerlof, G.A., & Yellen, J.L. (1996). New mothers, not married. *Brookings Review*, 14(4), 18–21.
- Alan Guttmacher Institute. (1976). *11 million teenagers: What can be done about the epidemic of adolescent pregnancies in the United States*. New York: Planned Parenthood of America.
- Allen, H. (1990). How to kill: Are Afrikan people the subjects of a genocidal plot? In A. Parfrey (Ed.), *Apocalypse culture* (rev. ed.). Portland, OR: Feral House.
- Amar, A.R., & Amar, D.A. (1996, May/June). Unlocking the jury box. *Policy Review*, pp. 38–44.
- Asher, H.B. (1976). *Causal modeling*. Sage University Series on Quantitative Applications in the Social Sciences, Series No. 07-003. Beverly Hills, CA: Sage.
- Associated Press. (1974, October 17). Disorganized crime just does not pay. *Evening Sun* (Baltimore), p. A1.
- Associated Press. (1988, March 10). Tripped up by a stub. *New York Daily News*, p. C9.
- Associated Press. (1995, November 24). AIDS study may blunt earlier optimism. *Washington Post*, p. A12.
- Bailey, J.M. (1989). A critique and reinterpretation of Gordon's IQ-commensurability property. *International Journal of Sociology and Social Policy*, 9, 64–73.
- Barton, P.E., & Jenkins, L. (1995). *Literacy and dependency: The literacy skills of welfare recipients in the United States*. Princeton: Educational Testing Service.
- Baur, E.J. (1960). Public opinion and the primary group. *American Sociological Review*, 25, 208–219.
- Bell, R.Q. (1968). A reinterpretation of the direction of effects in studies of socialization. *Psychological Review*, 75, 81–95.
- Berger, A.S., & Simon, W. (1974). Black families and the Moynihan Report: A research evaluation. *Social Problems*, 22, 145–161.
- Berlin, G., & Sum, A. (1988). *Toward a more perfect union: Basic skills, poor families, and our economic future*. New York: Ford Foundation.
- Biggs, J.B., & Collis, K.F. (1991). Multimodal learning and the quality of intelligent behavior. In H.A.H. Rowe (Ed.), *Intelligence: Reconceptualization and measurement*. Hillsdale, NJ: Erlbaum.
- Blau, P.M. (1960). Structural effects. *American Sociological Review*, 25, 178–193.
- Blumstein, A., Cohen, J., Roth, J.A., & Visher, C.A. (Eds.). (1986). *Criminal careers and "career criminals"* (Vol. 1). Washington, DC: National Academy Press.
- Bollinger, A.V., & Hoffmann, B. (1995, September 29). Cochran deals a deck of race cards. *New York Post*, pp. 4–5, 16.

- Bouchard, T.J., Jr., & McGue, M. (1981). Familial studies of intelligence: A review. *Science*, 212, 1055–1059.
- Bourgois, P. (1995). Poverty and unemployment cause crime. In *Crime and criminals: Opposing viewpoints*. San Diego, CA: Greenhaven Press.
- Brier, S.S., & Fienberg, S.E. (1980). Recent econometric modelling of crime and punishment: Support for the deterrence hypothesis? In S.E. Fienberg & A.J. Reiss, Jr. (Eds.), *Indicators of crime and criminal justice: Quantitative studies*. Washington, DC: U.S. Government Printing Office.
- Broder, D.S. (1995, October 11). Misjudging the jury system. *Washington Post*, p. A21.
- Bruni, F. (1996, September 30). Johnnie Cochran, author, takes his new cause to the pulpit. *New York Times*, p. B3.
- Bugliosi, V. (1996). *Outrage: The five reasons why O.J. Simpson got away with murder*. New York: Norton.
- Burt, C. (1948). *The young delinquent*. London: University of London Press. (Original work published 1925)
- Butler, P. (1995a, December). Black jurors: Right to acquit? *Harper's*, pp. 11–14, 16.
- Butler, P. (1995b, October 8). O.J. reckoning: Rage for a new justice. *Washington Post*, pp. C1–C2.
- Butler, D., Ray, A., & Gregory, L. (1995). *America's dumbest criminals*. Nashville: Rutledge Hill Press.
- Butterfield, F. (1996, March 2). A portrait of the detective in the "O.J. whirlpool." *New York Times*, pp. 1, 20.
- Cannon, L. (1995, October 8). It wasn't the jurors' fault. *Washington Post*, p. C7.
- Capaldi, D.M., Crosby, L., & Stoolmiller, M. (1996). Predicting the timing of first sexual intercourse for at-risk adolescent males. *Child Development*, 67, 344–359.
- Castaneda, R. (1995, September 29). Retarded man is freed in D.C. slaying. *Washington Post*, pp. A1, A16.
- Cattell, R.B., & Cattell, A.K.S. (1958). *Handbook for the individual or group: Culture Free Intelligence Test. Scale 2*. Champaign, IL: Institute for Personality and Ability Testing.
- Ceci, S.J., & Liker, J. (1986). Academic and nonacademic intelligence: An experimental separation. In R.J. Sternberg & R.K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world*. Cambridge: Cambridge University Press.
- Christenson, R. (1967). Projected percentage of U.S. population with criminal arrest and conviction records. In Institute for Defense Analysis, *Task force report: Science and technology* (Report to the President's Commission on Law Enforcement and the Administration of Justice). Washington, DC: U.S. Government Printing Office.
- Clark, Darden honor divides bar association. (1995, November 25). *The Sun* (Baltimore), p. 2A.
- Clark, K.B. (1965). *Dark ghetto: Dilemmas of social power*. New York: Harper & Row.
- Cochran, J.L., Jr. (1996). *Journey to justice*. New York: Ballantine Books.
- Cohen, J. (1959). The factorial structure of the WISC at ages 7-6, 10-6, and 13-6. *Journal of Consulting Psychology*, 23, 285–299.
- Cohen, J. (1986). Appendix B. Research on criminal careers: Individual frequency rates and offense seriousness. In A. Blumstein, J. Cohen, J.A. Roth, & C.A. Visher (Eds.), *Criminal careers and "career criminals"* (Vol. 1). Washington, DC: National Academy Press.
- Cohen, R. (1995, October 4). Reaction to O.J. verdict is a tale of two Americas. *New York Post*, p. 25.
- Cohn, N. (1966). *Warrant for genocide: The myth of the Jewish world conspiracy and the Protocols of the Elders of Zion*. New York: Harper & Row.
- Cooper, H. (1996, January 12). A question of justice: Do prosecutors target minority politicians? *Wall Street Journal*, pp. A1, A4.
- Crawford, S. (1994, May 8). Farrakhan upsets blacks in donor field. *Philadelphia Inquirer*, p. A15.
- Cross, J.F., & Cross, J. (1971). Age, sex, race, and the perception of facial beauty. *Developmental Psychology*, 5, 433–439.

- Crossen, C. (1994). *Tainted truth: The manipulation of fact in America*. New York: Simon & Schuster.
- Cunningham, M.R., Roberts, A.R., Barbee, A.P., Druen, P.B., & Wu, C. (1995). "Their ideas of beauty are, on the whole, the same as ours": Consistency and variability in the cross-cultural perception of female physical attractiveness. *Journal of Personality and Social Psychology*, 68, 261–279.
- Darden, C. (1996). *In contempt*. New York: ReganBooks.
- Davidson, J. (1996, October). Targets for scrutiny. *Emerge*, pp. 38–42, 44.
- Davis, A. (1943). Child training and social class. In R.G. Barker, J.S. Kounin, & H.F. Wright (Eds.), *Child behavior and development*. New York: McGraw-Hill.
- Dearden, K.A., Hale, C.B., & Woolley, T. (1995). The antecedents of teen fatherhood: A retrospective case-control study of Great Britain youth. *American Journal of Public Health*, 85, 551–554.
- Dégh, L. (1994). *American folklore and the mass media*. Bloomington: Indiana University Press.
- DeParle, J. (1990, October 29). Talk grows of government being out to get blacks. *New York Times*, p. B6.
- Dershowitz, A.M. (1996). *Reasonable doubts: The O.J. Simpson case and the criminal justice system*. New York: Simon & Schuster.
- Deutsch, L. (1995, September 29). Simpson defense urges a vote against racism. *The Sun* (Baltimore), pp. 1A, 10A.
- DiMascio, W.M. (1995). *Seeking justice: Crime and punishment in America*. New York: Edna McConnell Clark Foundation.
- Dixon, N. (1976). *On the psychology of military incompetence*. New York: Basic Books.
- Dow, D.R. (1995, November 10). Our jury system may have seen its day [Letter to the editor]. *New York Times*, p. A32.
- Dowd, M. (1995, October 5). O.J. as metaphor. *New York Times*, p. A29.
- D'Souza, D. (1995). *The end of racism: Principles for a multiracial society*. New York: Free Press.
- Duke, S. (1995, October 10). In Federal Court too [Letter to the editor]. *New York Times*, p. A22.
- Duncan, O.D. (1961a). A socioeconomic index for all occupations. In A.J. Reiss, Jr. (Ed.), *Occupations and social status*. New York: Free Press.
- Duncan, O.D. (1961b). Properties and characteristics of the socioeconomic index. In A.J. Reiss, Jr. (Ed.), *Occupations and social status*. New York: Free Press.
- Duncan, O.D., Featherman, D.L., & Duncan, B. (1972). *Socioeconomic background and achievement*. New York: Seminar Press.
- Dyson, T. (Ed.). (1991). *Sexual behavior and networking: Anthropological and socio-cultural studies on the transmission of HIV*. Liege, Belgium: International Union for the Scientific Study of Population.
- East, P.L. (1996). The younger sisters of childbearing adolescents: Their attitudes, expectations, and behaviors. *Child Development*, 67, 267–282.
- Edgerton, R.B. (1993). *The cloak of competence* (Rev. ed.). Berkeley: University of California Press.
- Edsall, T.B., & Edsall, M.D. (1992). *Chain reaction: The impact of race, rights, and taxes on American politics*. New York: Norton.
- Egan, T. (1995, October 5). With spotlight shifted to them, some Simpson jurors talk freely. *New York Times*, pp. A1, B18.
- Einhorn, H.J., & Hogarth, R.M. (1986). Judging probable cause. *Psychological Bulletin*, 99, 3–19.
- Eisner, J.R. (1995, October 15). Is Powell the one who can take us across that gaping racial divide? *Philadelphia Inquirer*, p. C7.
- Evidence points to deceit by Brawley. (1988, September 27). *New York Times*, pp. A1, B4–B5.
- Excerpts from final rebuttal arguments by prosecutors in Simpson case. (1995, September 30). *New York Times*, p. 8.
- False lessons of the Simpson trial. (1995, October 5). *New York Times*, p. A28.
- Fischer, C.S., Hout, M., Jankowski, M.S., Lucas, S.R., Swidler, A., Voss, K. (1996). *Inequality by design: Cracking the bell curve myth*. Princeton: Princeton University Press.

- Fiske, J. (1996). *Media matters: Everyday culture and political change* (Rev. ed.). Minneapolis: University of Minnesota Press.
- Fitzgerald, J.M., & Mellor, S. (1988). How do people think about intelligence? *Multivariate Behavioral Research*, 23, 143–157.
- Fleming, J.E. (1976). *The lengthening shadow of slavery: Historical justification for affirmative action for blacks in higher education*. Washington, DC: Howard University Press.
- Fletcher, M.A. (1996, October 4). Conspiracy theories can often ring true. *Washington Post*, pp. A1, A18.
- Ford, M.E. (1986). For all practical purposes: Criteria for defining and evaluating practical intelligence. In R.J. Sternberg & R.K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world*. Cambridge: Cambridge University Press.
- Fraser, S. (Ed.). (1995). *The bell curve wars: Race, intelligence, and the future of America*. New York: Basic Books.
- Freud, S. (1965). *The psychopathology of everyday life*. New York: Norton. (Original work published 1901)
- Gallagher, M. (1996). *The abolition of marriage*. Washington, DC: Regnery.
- Gardner, E.S. (1958). Introduction. In N.F. Leopold, *Life plus 99 years*. Garden City, NY: Doubleday.
- Gardner, H. (1985). *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.
- Glueck, S., & Glueck, E. (1950). *Unraveling juvenile delinquency*. Cambridge, MA: Harvard University Press.
- Goldberg, H.M. (1996). *The prosecution responds*. Secaucus, NJ: Birch Lane Press.
- Golden, T. (1996a, October 21). Pivotal figures of newspaper series may be only bit players. *New York Times*, p. A14.
- Golden, T. (1996b, October 21). Tale of C.I.A. and drugs has life of its own. *New York Times*, pp. A1, A14.
- Goodnow, J.J. (1986). A social view of intelligence. In R.J. Sternberg & D.K. Detterman (Eds.), *What is intelligence?* Norwood, NJ: Ablex.
- Gordon, R.A. (1967). Issues in the ecological study of delinquency. *American Sociological Review*, 32, 927–944.
- Gordon, R.A. (1973). An explicit estimation of the prevalence of commitment to a training school, to age 18, by race and by sex. *Journal of the American Statistical Association*, 68, 547–553.
- Gordon, R.A. (1976). Prevalence: The rare datum in delinquency measurement and its implications for the theory of delinquency. In M.W. Klein (Ed.), *The juvenile justice system*. Beverly Hills, CA: Sage.
- Gordon, R.A. (1977). A critique of the evaluation of Patuxent Institution, with particular attention to the issues of dangerousness and recidivism. *Bulletin of the American Academy of Psychiatry and the Law*, 5, 210–255.
- Gordon, R.A. (1980a). Examining labelling theory: The case of mental retardation. In W.R. Gove (Ed.), *The labelling of deviance* (2nd ed.). Beverly Hills, CA: Sage. (Original work published 1975)
- Gordon, R.A. (1980b). Implications of valid (and stubborn) IQ differences: An unstatesmanlike view. *Behavioral and Brain Sciences*, 3, 343–344.
- Gordon, R.A. (1980c). Labelling theory, mental retardation, and public policy: *Larry P.* and other developments since 1974. In W.R. Gove (Ed.), *The labelling of deviance* (2nd ed.). Beverly Hills, CA: Sage.
- Gordon, R.A. (1980d). Research on IQ, race, and delinquency: Taboo or not taboo? In E. Sagarin (Ed.), *Taboos in criminology*. Beverly Hills, CA: Sage.
- Gordon, R.A. (1982). Preventive sentencing and the dangerous offender. *British Journal of Criminology*, 22, 285–314.
- Gordon, R.A. (1984). Digits backward and the Mercer-Kamin law. In C.R. Reynolds & R.T. Brown (Eds.), *Perspectives on bias in mental testing*. New York: Plenum.

- Gordon, R.A. (1986). Scientific justification and the race-IQ-delinquency model. In T.F. Hartnagel & R.A. Silverman (Eds.), *Critique and explanation: Essays in honor of Gwynne Nettler*. New Brunswick, NJ: Transaction Books.
- Gordon, R.A. (1987). SES versus IQ in the race-IQ-delinquency model. *International Journal of Sociology and Social Policy*, 7, 30-96.
- Gordon, R.A. (1988). Thunder from the left. *Academic Questions*, 1(3), 74-92.
- Gordon, R.A. (1995). The great disturbance about intelligence. *Planning for Higher Education*, 23(3), 19-28.
- Gordon, R.A., & Gleser, L.J. (1974). The estimation of the prevalence of delinquency: Two approaches and a correction of the literature. *Journal of Mathematical Sociology*, 3, 275-291.
- Gordon, R.A., Lewis, M.A., & Quigley, A.M. (1988). Can we count on muddling through the crisis in employment? *Journal of Vocational Behavior*, 33, 424-451.
- Gordon, R.A., & Rudert, E.E. (1979). Bad news concerning IQ tests. *Sociology of Education*, 52, 174-190.
- Gordon, R.A., Short, J.F., Jr., Cartwright, D.S., and Strodtbeck, F.L. (1963). Values and gang delinquency: A study of street-corner groups. *American Journal of Sociology*, 69, 109-128.
- Gottfredson, L.S. (1985). Education as a valid but fallible signal of worker quality: Reorienting an old debate about the functional basis of the occupational hierarchy. In A.C. Kerckhoff (Ed.), *Research in Sociology of Education and Socialization*, 5, 123-169.
- Gottfredson, L.S. (1986a). Societal consequences of the g factor in employment. *Journal of Vocational Behavior*, 29, 379-410.
- Gottfredson, L.S. (Ed.). (1986b). The g factor in employment [Special issue]. *Journal of Vocational Behavior*, 29(3).
- Gottfredson, L.S. (1994). The science and politics of race-norming. *American Psychologist*, 49, 955-963.
- Gottfredson, L.S. (1997). Why g matters: The complexity of everyday life. *Intelligence*, 24(1), 79-132.
- Gove, W.R. (1980). Postscript. In W.R. Gove (Ed.), *The labelling of deviance* (2nd ed.). Beverly Hills, CA: Sage.
- Gould, S.J. (1994, November 28). Curveball. *New Yorker*, pp. 139-149.
- Gouldner, A.W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, 25, 161-178.
- Grammer, K., & Thornhill, R. (1994). Human (*Homo sapiens*) facial attractiveness and sexual selection: The role of symmetry and averageness. *Journal of Comparative Psychology*, 108, 233-242.
- Green, B.F., Jr. (1978). In defense of measurement. *American Psychologist*, 33, 664-670.
- Greene, M.H. (1974). *Estimating the prevalence of heroin use in a community*. Special Action Office Monograph Series A, No. 4. Washington, DC: Executive Office of the President, Special Action Office for Drug Abuse Prevention.
- Greenberg, P. (1995, October 13). About those two nations . . . *Washington Times*, p. A20.
- Gresham, F.M., MacMillan, D.L., & Siperstein, G.N. (1995). Critical analysis of the 1992 AAMR definition: Implications for school psychology. *School Psychology Quarterly*, 10, 1-19.
- Haberman, C. (1995, November 26). Color-blind? Justice, maybe, but not juries. *New York Times*, p. 39.
- Hacker, A. (1992). *Two nations: Black and white, separate, hostile, unequal*. New York: Scribner's.
- Hallow, R.Z. (1995, August 31). Black and white answers on racism are elusive: Polls on Simpson case show schism. *Washington Times*, pp. A1, A8.
- Harrington, E. (1996, September/October). Conspiracy theories and paranoia: Notes from a mind-control conference. *Skeptical Inquirer*, pp. 35-42.
- Harris, J.F. (1995, October 11). Clinton "struck" by racial divide on Simpson case. *Washington Post*, p. A13.

- Hawk, J.A. (1970). Linearity of criterion–GATB aptitude relationships. *Measurement and Evaluation in Guidance*, 2, 249–251.
- Hawley, A.H. (1950). *Human ecology: A theory of community*. New York: Ronald Press.
- Heard, E. (1995, September 3). Slain man's survivors, supporters protest justice system. *The Sun* (Baltimore), p. 2C.
- Hentoff, N. (1995, October 10). In defense of the Simpson jury. *Washington Post*, p. A13.
- Hermann, P. (1995, July 31). Questions linger about Lee verdict. *The Sun* (Baltimore), pp. 1B, 3B.
- Hermann, P., & Myers, M. (1996, March 30). Suspect charged in shooting at Rotunda Giant. *The Sun* (Baltimore), pp. 1B–2B.
- Hernandez, D.J. (1993). *America's children: Resources from family, government, and the economy*. New York: Russell Sage Foundation.
- Herrnstein, R.J., & Murray, C. (1994). *The bell curve: Intelligence and class in American life*. New York: Free Press.
- Hirschi, T. (1969). *Causes of delinquency*. Berkeley: University of California Press.
- Hirschi, T., & Hindelang, M.J. (1977). Intelligence and delinquency: A revisionist review. *American Sociological Review*, 42, 571–587.
- Hispanic juror speaks out on Simpson verdict. (1995, October 10). *The Sun* (Baltimore), p. 5A.
- Hochschild, J.L. (1995). *Facing up to the American dream: Race, class, and the soul of the nation*. Princeton: Princeton University Press.
- Hofstadter, R. (1965). *The paranoid style in American politics*. New York: Knopf.
- Holden, B.A., Cohen, L.P., & de Lissar, E. (1995, October 4). Race seems to play an increasing role in many jury verdicts. *Wall Street Journal*, pp. A1, A5.
- Holmberg, S.D. (1996). The estimated prevalence and incidence of HIV in 96 large U.S. metropolitan areas. *American Journal of Public Health*, 86, 642–654.
- Hopkins, E. (1995, October 8). After the verdict, solidarity in the cell block. *Washington Post*, pp. C1–C2.
- Houts, P.L. (1977). A conversation with Banesh Hoffmann. In P.L. Houts (Ed.), *The myth of measurability*. New York: Hart.
- Howe, M.J.A. (1990). *The origins of exceptional abilities*. Cambridge, MA: Basil Blackwell.
- Huber, J., & Chalfant, H.P. (Eds.). (1974). *The sociology of American poverty*. Cambridge, MA: Schenkman Publishing Company.
- Humphreys, L.G. (1992). Commentary: What both critics and users of ability tests need to know. *Psychological Science*, 3, 271–274.
- Hunter, J.E., & Schmidt, F.L. (1982). Fitting people to jobs: The impact of personnel selection on national productivity. In M.D. Dunnette & E.A. Fleishman (Eds.), *Human performance and productivity: Human capability assessment*. Hillsdale, NJ: Erlbaum.
- Hunter, J.E., Schmidt, F.L., & Rauschenberger, J. (1984). Methodological, statistical, and ethical issues in the study of bias in psychological tests. In C.R. Reynolds & R.T. Brown (Eds.), *Perspectives on bias in mental testing*. New York: Plenum.
- Hurley, R.L. (1969). *Poverty and mental retardation: A causal relationship*. New York: Random House.
- Hyman, H.H., & Sheatsley, P.B. (1947). Some reasons why information campaigns fail. *Public Opinion Quarterly*, 11, 412–423.
- Irons, W. (1996). In our own self image: The evolution of morality, deception, and religion. *Skeptic*, 4, 50–61.
- Iskrant, A.P., & Joliet, P.V. (1968). *Accidents and homicide*. Cambridge, MA: Harvard University Press.
- Jacoby, R., & Glaubergerman, N. (Eds.). (1995). *The bell curve debate: History, documents, opinions*. New York: Times Books.

- James, M., & West, N.P. (1995, August 2). Jurors defend verdict in slaying of student. *The Sun* (Baltimore), pp. 1B, 3B.
- Janes, R.C. (1995, December 26). The myth of impartial juries [Letter to the editor]. *Wall Street Journal*, p. A7.
- Jaynes, G.D., & Williams, R.M., Jr. (Eds.). (1989). *A common destiny: Blacks and American society*. Washington, DC: National Academy Press.
- Jensen, A.R. (1980a). *Bias in mental testing*. New York: Free Press.
- Jensen, A.R. (1980b). Chronometric analysis of intelligence. *Journal of Social and Biological Structures*, 3, 103–122.
- Jensen, A.R. (1981). *Straight talk about mental tests*. New York: Free Press.
- Jensen, A.R. (1984). Test validity: *g* versus the specificity doctrine. *Journal of Social and Biological Structures*, 7, 93–118.
- Jensen, A.R. (1985). The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, 8, 193–263.
- Jensen, A.R. (1986a). *g*: Artifact or reality? *Journal of Vocational Behavior*, 29, 301–331.
- Jensen, A.R. (1986b). Intelligence: "Definition," measurement, and future research. In R.J. Sternberg & D.K. Detterman (Eds.), *What is intelligence?* Norwood, NJ: Ablex.
- Jensen, A.R. (1993a). Psychometric *g* and achievement. In B.R. Gifford (Ed.), *Policy perspectives on educational testing*. Boston: Kluwer.
- Jensen, A.R. (1993b). Spearman's *g*: Links between psychometrics and biology. *Annals of the New York Academy of Sciences* (Brain Mechanisms), 702, 103–129.
- Jeter, J. (1995, October 12). The O.J. effect: Juror in P.G. case sends his own message. *Washington Post*, pp. B1, B7.
- Jones, D., & Hill, K. (1993). Criteria of facial attractiveness in five populations. *Human Nature*, 4, 271–296.
- Juror #6. (1995, October 6). In *Dateline*. New York: NBC.
- Kahlenberg, R.D. (1996). *The remedy: Class, race, and affirmative action*. New York: Basic Books.
- Kandel, D.B. (1978). Homophily, selection, and socialization in adolescent friendships. *American Journal of Sociology*, 84, 427–434.
- Kaplan, E.H., & Brandeau, M.L. (Eds.). (1994). *Modeling the AIDS epidemic: Planning, policy, and prediction*. New York: Raven.
- Karon, J.M., Rosenberg, P.S., McQuillan, G., Khare, M., Gwinn, M., & Petersen, L.R. (1996). Prevalence of HIV infection in the United States, 1984 to 1992. *Journal of the American Medical Association*, 276, 126–131.
- Katz, J. (1996, October 20). Tracking the genesis of the crack trade in L.A. *Los Angeles Times*, pp. A1, A14–A15.
- Katz, E., & Lazarsfeld, P.F. (1955). *Personal influence: The part played by people in the flow of mass communications*. Glencoe, IL: Free Press.
- Kaufman, A.S. (1975). Factor analysis of the WISC-R at 11 age levels between 6½ and 16½ years. *Journal of Consulting and Clinical Psychology*, 43, 135–147.
- Kaufman, A.S., & Doppelt, J.E. (1976). Analysis of the WISC-R standardization data in terms of the stratification process. *Child Development*, 47, 165–171.
- Kelley, T.L. (1947). *Fundamentals of statistics*. Cambridge, MA: Harvard University Press.
- Kennedy, R. (1994, September 30). The angry juror. *Wall Street Journal*, p. A12.
- Kennedy, W.A., Van De Riet, V., & White, J.C., Jr. (1963). A normative sample of intelligence and achievement of Negro elementary school children in the Southeastern United States. *Mono-graphs of the Society for Research in Child Development*, 28(6, Serial No. 90).

- Kincheloe, J.L., Steinberg, S.R., & Gresson, A.D., III. (1996). *Measured lies: The bell curve examined*. New York: St. Martin's.
- Kinder, D.R., & Sanders, L.M. (1996). *Divided by color; Racial politics and democratic ideals*. Chicago: University of Chicago Press.
- King, C.I. (1995, October 7). The dead end of racial separation. *Washington Post*, p. A29.
- King, C.I. (1996, July 13). Tia Mitchell's murder. *Washington Post*, p. A19.
- Kleinbaum, D.G., Kupper, L.L., & Morgenstern, H. (1982). *Epidemiological research: Principles and quantitative methods*. Belmont, CA: Wadsworth.
- Knupper, G. (1947). Portrait of the underdog. *Public Opinion Quarterly*, 11, 101–114.
- Kramon, J.M. (1995, December 3). O.J.'s jury did the right thing. *The Sun* (Baltimore), pp. 1J, 6J.
- Kretzschmar, M., Reinking, D.P., Brouwers, H., van Zessen, G., & Jager, J.C. (1994). Network models: From paradigm to mathematical tool. In E.H. Kaplan & M.L. Brandeau (Eds.), *Modeling the AIDS epidemic: Planning, policy, and prediction*. New York: Raven.
- Kropotkin, P. (1989). *Mutual aid: A factor of evolution*. New York: Black Rose Books. (Original work published 1902)
- Kurtz, H. (1996, October 2). Running with the CIA story; Reporter says series didn't go as far as readers took it. *Washington Post*, pp. B1, B8.
- Ladd, E.C. (1996, February/March). Moving to an America beyond race. *The Public Perspective*, pp. 36–42.
- Lambert, B. (1988, October 13). Panel warns of increase in AIDS cases. *New York Times*, pp. B1, B5.
- Langan, P.A. (1991). America's soaring prison population. *Science*, 251, 1568–1573.
- Langlois, J.H., Roggman, L.A., Casey, R.J., Ritter, J.M., Rieser-Danner, L.A., & Jenkins, V.Y. (1987). Infant preferences for attractive faces: Rudiments of a stereotype? *Developmental Psychology*, 23, 363–369.
- Laub, J.H. (1983). Urbanism, race, and crime. *Journal of Research in Crime and Delinquency*, 20, 183–198.
- Laughlin, P.R., & Johnson, H.H. (1966). Group and individual performance on a complementary task as a function of initial ability level. *Journal of Experimental Social Psychology*, 2, 407–414.
- Leape, L.L. (1994). Error in medicine. *Journal of the American Medical Association*, 272, 1851–1857.
- Leary, W. (1995, June 4). AIDS and tuberculosis epidemics rise in deadly combination. *New York Times*, p.8.
- Leopold, N.F. (1958). *Life plus 99 years*. Garden City, NY: Doubleday.
- Lerner, R. (1996). Acquittal rates by race for state felonies. In G.A. Reynolds (Ed.), *Race and the criminal justice system: How race affects jury trials*. Washington, DC: Center for Equal Opportunity.
- Levitan, S.A. (1969). *The Great Society's poor law: A new approach to poverty*. Baltimore: Johns Hopkins University Press.
- Lewis, J. (1995, July). Crime as a cause of poverty. *Atlantic Monthly*, p. 49.
- Lewis, O. (1966, October). The culture of poverty. *Scientific American*, pp. 19–25.
- Lewis, T. (1995). Carroll in wonderland [Review of the book *Lewis Carroll: A biography*]. *Wall Street Journal*, p. A18.
- Light, R.J., & Pillemer, D.B. (1984). *Summing up: The science of reviewing research*. Cambridge, MA: Harvard University Press.
- Linton, R. (1936). *The study of man*. New York: Appleton-Century-Crofts.
- Littlepage, G.E., Schmidt, G.W., Whisler, E.W., & Frost, A.G. (1995). An input–process–output analysis of influence and performance in problem-solving groups. *Journal of Personality and Social Psychology*, 69, 877–889.
- Littwin, M. (1995, October 4). The great divide. *The Sun* (Baltimore), pp. 1E, 5E.

- Loftin, C. (1980). Alternative estimates of the impact of certainty and severity of punishment on levels of homicide in American states. In S.E. Fienberg & A.J. Reiss, Jr. (Eds.), *Indicators of crime and criminal justice: Quantitative studies*. Washington, DC: U.S. Government Printing Office.
- Lubinski, D., & Humphreys, L.G. (1997). Incorporating general intelligence into epidemiology and the social sciences. *Intelligence*, 24(1), 159–201.
- Lykken, D.T. (1995). *The antisocial personalities*. Hillsdale, NJ: Erlbaum.
- Lynn, R. (1996). Racial and ethnic differences in intelligence in the United States on the Differential Ability Scale. *Personality and Individual Differences*, 20, 271–273.
- Magnet, M. (1993). *The dream and the nightmare*. New York: William Morrow.
- Maller, J.B. (1933). Vital indices and their relation to psychological and social factors. *Human Biology*, 5, 94–121.
- Margolick, D. (1995, September 29). With tale of racism and error, Simpson lawyers seek acquittal. *New York Times*, pp. A1, A22.
- Marsden, P. (1988). Homogeneity in confiding relations. *Social Networks*, 10, 57–76.
- Mauer, M., & Huling, T. (1995). *Young black Americans and the criminal justice system: Five years later*. Washington, DC: The Sentencing Project.
- McAneny, L. (1995, January). It was a very bad year: belief in Hell and the Devil are on the rise. *Gallup Poll Monthly*, pp. 14–17.
- McCabe, D.L., & Trevino, L.K. (1996, January/February). What we know about cheating in college: Longitudinal trends and recent developments. *Change*, pp. 29–33.
- McCord, W., & McCord, J. (1959). *Origins of crime: A new evaluation of the Cambridge–Somerville study*. New York: Columbia University Press.
- McFadden, R.D. (1988a, October 29). Brawley case: Public's conflicting views. *New York Times*, pp. 1, 33.
- McFadden, R.D. (1988b, October 7). Brawley made up story of assault, grand jury finds. *New York Times*, pp. A1, B4.
- McFarland, R.A., & Moore, R.C. (1961). Youth and the automobile. In E. Ginzberg (Ed.), *Values and ideals of youth*. New York: Columbia University Press.
- McLellan, D. (1978). *Karl Marx*. New York: Penguin.
- McPherson, M., & Smith-Lovin, L. (1987). Homophily in voluntary organizations. *American Sociological Review*, 52, 370–379.
- Meier, B. (1996, June 11). Infected hemophiliacs and tainted lawsuits. *New York Times*, pp. D1, D21.
- Mercer, J.R. (1973). *Labelling the retarded*. Berkeley: University of California Press.
- Mercer, J.R. (1979). *SOMPA: Technical manual*. New York: Psychological Corporation.
- Mercer, J.R. (1984). What is a racially and culturally discriminating test? In C.R. Reynolds & R.T. Brown (Eds.), *Perspectives on bias in mental testing*. New York: Plenum.
- Merton, R.K. (1957). *Social theory and social structure* (Rev. ed.). Glencoe, IL: Free Press.
- Miele, F. (1979). Cultural bias in the WISC. *Intelligence*, 3, 149–164.
- Miller, W.B. (1958). Lower class culture as a generating milieu of gang delinquency. *Journal of Social Issues*, 14(3), 5–19.
- Milloy, C. (1995, October 8). In white riot, it's smolder, baby, smolder. *Washington Post*, p. B1.
- Moffitt, T.E., & Silva, P.A. (1988). IQ and delinquency: A direct test of the differential detection hypothesis. *Journal of Abnormal Psychology*, 97, 330–333.
- Moffitt, T.E., Gabrielli, W.F., Mednick, S.A., & Schulsinger, F. (1981). Socioeconomic status, IQ, and delinquency. *Journal of Abnormal Psychology*, 90, 152–156.
- Moore, D.W. (1995, January). Order of presentation can affect public's responses. *Gallup Poll Monthly*, pp. 18–23.

- Moore, D.W., & Saad, L. (1995, October). No immediate signs that Simpson trial intensified racial animosity. *Gallup Poll Monthly*, pp. 2–9.
- Morin, R. (1995, October 8). Poll reflects division over Simpson case. *Washington Post*, pp. A31, A34.
- Moynihan, D.P. (1965). *The Negro family: The case for national action* (U.S. Department of Labor). Washington, DC: U.S. Government Printing Office.
- Moynihan, D.P. (1993). Defining deviancy down. *American Scholar*, 62, 17–30.
- Murray, C. (1984). *Losing ground: American social policy 1950–1980*. New York: Basic Books.
- Myers, S.L., Jr. (1990, March 29). Conspiracy and black political leaders. *Evening Sun* (Baltimore), p. A19.
- National Science Board. (1996). *Science and engineering indicators—1996*. Washington, DC: U.S. Government Printing Office.
- Neal, D.A., & Johnson, W.R. (1996). The role of premarket factors in black–white wage differences. *Journal of Political Economy*, 104, 869–895.
- Nevo, B., & Khader, A.M.B. (1995). Cross-cultural, gender, and age differences in Singaporean mothers' conceptions of children's intelligence. *Journal of Social Psychology*, 135, 509–517.
- Noble, K.B. (1996, April 2). District attorney investigates accusations of jury tampering in Simpson case. *New York Times*, p. A12.
- Norbom, M.A. (1995, October 17). O.J. killed 'em. *Globe*, pp. 4–5, 8.
- Oakland, T. (1983). Joint use of adaptive behavior and IQ to predict achievement. *Journal of Consulting and Clinical Psychology*, 51, 298–301.
- Oliver's world. (1995, December 24). *New York Times*, p. 8E.
- Orfield, G., & Ashkinaze, C. (1991). *The closing door: Conservative policy and black opportunity*. Chicago: University of Chicago Press.
- Orshansky, M. (1969, February). How poverty is measured. *Monthly Labor Review*, 92, 37–41.
- Orubuloye, I.O., Caldwell, J.C., Caldwell, P., & Santow, G. (Eds.). (1994). *Sexual networking and AIDS in sub-Saharan Africa: Behavioral research and the social context*. Canberra: Australian National University Printing Service.
- Orwell, G. (1968). James Burnham and the managerial revolution. In S. Orwell and I. Angus (Eds.), *The collected essays, journalism and letters of George Orwell* (Vol. 4). (Original work published 1946)
- Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). *The measurement of meaning*. Urbana: University of Illinois Press.
- Osgood, C.E., Ware, E.E., & Morris, C. (1961). Analysis of the connotative meaning of a variety of human values as expressed by American college students. *Journal of Abnormal and Social Psychology*, 62, 63–73.
- O'Toole, B.I. (1990). Intelligence and behaviour and motor vehicle accident mortality. *Accident Analysis & Prevention*, 22, 211–221.
- Patton, R. (1996, March). [Letter to the editor]. *Commentary*, p. 12.
- Paulos, J.A. (1995, October 15). What "statisticide" tells about the Simpson case. *Philadelphia Inquirer*, p. C7.
- Peter, L.J., & Hull, R. (1969). *The Peter principle*. New York: William Morrow.
- Phillipson, T.J., & Posner, R.A. (1993). *Private choices and public health: The AIDS epidemic in an economic perspective*. Cambridge, MA: Harvard University Press.
- Plomin, R., & Petrill, S.A. (1997). Genetics and intelligence: What's new? *Intelligence*, 24(1), 53–77.
- Plotnick, R.D., & Skidmore, F. (1975). *Progress against poverty: A review of the 1964–1974 decade*. New York: Academic.
- Purnick, J. (1995, October 12). Finding out how to weed out the skeptics from juries. *New York Times*, pp. B1, B5.
- Quinney, R. (1970). *The social reality of crime*. Boston: Little, Brown.

- Race cards and rebuttals. (1995, September 30). *New York Times*, p. 18.
- Rainwater, L., & Yancey, W.L. (1967). *The Moynihan Report and the politics of controversy*. Cambridge, MA: M.I.T. Press.
- Ranchor, A.V., Bouma, J., & Sanderman, R. (1996). Vulnerability and social class: Differential patterns of personality and social support over the social classes. *Personality and Individual Differences*, 20, 229–237.
- Raspberry, W. (1995a, October 23). A nation of poor listeners. *Washington Post*, p. A21.
- Raspberry, W. (1995b, October 5). Just what was the “race card”? *Washington Post*, p. A23.
- Raspberry, W. (1996, September 23). The crack story: Who’s buying it? *Washington Post*, p. A19.
- Reibstein, L. (1995, October 9). Judgment day. *Time*, pp. 26–32.
- Reiss, A.J., Jr. (1986). Co-offender influences on criminal careers. In A. Blumstein, J. Cohen, J.A. Roth, & C.A. Visher (Eds.), *Criminal careers and “career criminals”* (Vol. 2). Washington, DC: National Academy Press.
- Reynolds, C.R., Chastain, R.L., Kaufman, A.S., & McLean, J.E. (1987). Demographic characteristics and IQ among adults: Analysis of the WAIS-R standardization sample as a function of the stratification variables. *Journal of School Psychology*, 25, 323–342.
- Riley, J.L. (1995, November 22). Should a jury verdict be unanimous? *Wall Street Journal*, p. A11.
- Robins, L.N. (1966). *Deviant children grown up: A sociological and psychiatric study of sociopathic personality*. Baltimore: Williams & Wilkins.
- Rosen, L., & Neilson, K. (1978). The broken home and delinquency. In L.D. Savitz & N. Johnston (Eds.), *Crime in society*. New York: Wiley.
- Rosenberg, H.L. (Producer). (1996, March 31). The letter. In *60 Minutes*. New York: CBS.
- Rosenberg, P.S. (1995). Scope of the AIDS epidemic in the United States. *Science*, 270, 1372–1375.
- Ross, S. (Producer). (1996, October 8). *Mark Fuhrman speaks*. In PrimeTime Live. New York: ABC.
- Roth, B.M. (1994). *Prescription for failure: Race relations in the age of social science*. New Brunswick, NJ: Transaction.
- Rothwax, H.J. (1996). *Guilty: The collapse of the criminal justice system*. New York: Random House.
- Rowan, C.T. (1995, August 4). No racial monolith on O.J. *The Sun* (Baltimore), p. 15A.
- Rowe, D.C., & Rodgers, J.L. (1994). A social contagion model of adolescent sexual behavior: Explaining race differences. *Social Biology*, 41, 1–18.
- Rowe, D.C., Rodgers, J.L., & Meseck-Bushey, S. (1989). An “epidemic” model of sexual intercourse prevalences for black and white adolescents. *Social Biology*, 36, 127–145.
- Rowe, D.C., Vazsonyi, A.T., & Flannery, D.J. (1994). No more than skin deep: Ethnic and racial similarity in developmental process. *Psychological Review*, 101, 396–413.
- Rubenstein, D., & Horowitz, C.F. (1995). Poverty and unemployment do not cause crime. In *Crime and criminals: Opposing viewpoints*. San Diego, CA: Greenhaven Press.
- Ruggles, P. (1990). *Drawing the line: Alternative poverty measures and their implications for public policy*. Washington, DC: Urban Institute Press.
- Ryan, W. (1971). *Blaming the victim*. New York: Vintage Books.
- Sachs, A. (1995, October 9). Drawing lines and lessons. *Time*, pp. 36–37.
- Safire, W. (1995, October 12). After the aftermath. *New York Times*, p. A23.
- Sampson, R.J., & Laub, J.H. (1993). *Crime in the making: Pathways and turning points through life*. Cambridge, MA: Harvard University Press.
- Sanders, C.E., Lubinski, D., & Benbow, C.P. (1995). Does the Defining Issues Test measure psychological phenomena distinct from verbal ability?: An examination of Lykken’s query. *Journal of Personality and Social Psychology*, 69, 498–504.
- Schmitz, S.H., & Castillo-Chavez, C. (1994). Parameter estimation in nonclosed social networks related to dynamics of sexually transmitted diseases. In E.H. Kaplan & M.L. Brandeau (Eds.), *Modeling the AIDS epidemic: Planning, policy, and prediction*. New York: Raven.

- Schofield, J.W. (1982). *Black and white in school: Trust, tension, or tolerance?* New York: Praeger.
- Senders, J.W., & Moray, N.P. (1991). *Human error: Cause, prediction, and reduction*. Hillsdale, NJ: Erlbaum.
- Shapiro, R.L. (1996). *The search for justice: A defense attorney's brief on the O.J. Simpson case*. New York: Warner Books.
- Sharf, J.C. (1988). Litigating personnel measurement policy. *Journal of Vocational Behavior*, 33, 235–271.
- Shoemaker, D.J. (1996). *Theories of delinquency: An examination of explanations of delinquent behavior* (3rd ed.). New York: Oxford University Press.
- Simonds, J.F., & Kashani, J. (1979). Drug abuse and criminal behavior in delinquent boys committed to a training school. *American Journal of Psychiatry*, 136, 1444–1448.
- Simonton, D.K. (1985). Intelligence and personal influence in groups: Four nonlinear models. *Psychological Review*, 92, 532–547.
- Simpson trial & trivia. (1995, October 16). *U.S. News and World Report*, pp. 42–43.
- Smith, S. (1995, October 11). [Letter to the editor]. *Washington Post*, p. A20.
- Snyderman, M., & Rothman, S. (1988). *The IQ controversy: The media and public policy*. New Brunswick, NJ: Transaction.
- Spizzerri, C. (1996, March 22). Two-thirds of new AIDS cases strike blacks. *The Review* (University of Delaware), p. A2.
- Stanley, J.C. (1971). Reliability. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed.). Washington, DC: American Council on Education.
- Sternberg, R.J. (1987). “Gee, there’s more than g!”: A critique of Arthur Jensen’s views on intelligence. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy*. New York: Falmer.
- Sternberg, R.J. (1988). *The triarchic mind: A new theory of human intelligence*. New York: Viking.
- Sternberg, R.J. (1996, March). Myths, countermyths, and truths about intelligence. *Educational Researcher*, pp. 11–16.
- Sternberg, R.J., & Wagner, R.K. (Eds.). (1986). *Practical intelligence: Nature and origins of competence in the everyday world*. Cambridge: Cambridge University Press.
- Streisand, B. (1995, October 16). The verdict’s aftermath. *U.S. News and World Report*, pp. 34–35, 38, 40.
- Summary. (1990, December). Workshop on family structure and ethnicity. Sponsored by the Program on Human Development and Criminal Behavior, Harvard School of Public Health, New Orleans, LA.
- Suro, R., & Pincus, W. (1996, October 4). The CIA and crack: Evidence is lacking of an alleged plot. *Washington Post*, pp. A1, A19.
- Sutherland, E.H. (1931). Mental deficiency and crime. In K. Young (Ed.), *Social attitudes*. New York: Holt.
- Sutherland, E.H. (1939). *Principles of criminology* (3rd ed.). Philadelphia: Lippincott.
- Tarantola, D., Mann, J., Mantel, C., & Cameron, C. (1994). Projecting the course of the HIV/AIDS pandemic and the cost of adult AIDS care in the world. In E.H. Kaplan & M.L. Brandeau (Eds.), *Modeling the AIDS epidemic: Planning, policy, and prediction*. New York: Raven.
- Taylor, H. (1995, February/March). Horses for courses: How different countries measure public opinion in very different ways. *The Public Perspective*, pp. 3–7.
- Taylor, J. (1992). *Paved with good intentions: The failure of race relations in contemporary America*. New York: Carroll & Graf.
- Telephone helps solve survey problems. (1978). *ISR Newsletter*, 6(1), 3.

- Terman, L.M. (1956). *Manual for the Concept Mastery Test*. New York: Psychological Corporation.
- Terman, L.M., & Merrill, M.A. (1960). *Stanford-Binet Intelligence Scale: Manual for the third revision, Form L-M*. Boston: Houghton Mifflin.
- Thernstrom, A. (1995, October 12). Two nations, separate and hostile? *New York Times*, p. A23.
- Thomas, A.P. (1994). *Crime and the sacking of America: The roots of chaos*. McLean, VA: Brassey's.
- Tittle, C.R. (1980). Labelling and crime: An empirical evaluation. In W.R. Gove (Ed.), *The labelling of deviance* (2nd ed.). Beverly Hills, CA: Sage. (Original work published 1975)
- Tonry, M. (1995). *Malign neglect: Race, crime, and punishment in America*. New York: Oxford University Press.
- Toobin, J. (1996). *The run of his life: The people v. O.J. Simpson*. New York: Random House.
- Tracy, P.E., Wolfgang, M.E., & Figlio, R.M. (1985). *Delinquency in two birth cohorts*. Chicago: University of Chicago Press.
- Tubman, J.G., Windle, M., & Windle, R.C. (1996). The onset and cross-temporal patterning of sexual intercourse in middle adolescence: Prospective relations with behavioral and emotional problems. *Child Development*, 67, 327–343.
- Turner, P.A. (1993). *I heard it through the grapevine: Rumor in African-American culture*. Berkeley: University of California Press.
- U.S. Bureau of the Census. (1979). *The social and economic status of the black population in the United States: An historical view, 1790–1978*. (Current Population Reports: Special Studies: Series P-23, No. 80). Washington, DC: U.S. Government Printing Office.
- U.S. Bureau of the Census. (1992). *Educational attainment in the United States: March 1991 and 1990*. (Current Population Reports, Series P-20, No. 462). Washington, DC: U.S. Government Printing Office.
- U.S. Bureau of the Census. (1993). *Poverty in the United States: 1992*. (Current Population Reports, Series P60-185). Washington, DC: U.S. Government Printing Office.
- U.S. Bureau of the Census. (1994). *Household and family characteristics: March 1993*. (Current Population Reports, P20-477). Washington, DC: U.S. Government Printing Office.
- Visher, C.A., & Roth, J.A. (1986). Appendix A. Participation in criminal careers. In A. Blumstein, J. Cohen, J.A. Roth, & C.A. Visher (Eds.), *Criminal careers and "career criminals"* (Vol. 1). Washington, DC: National Academy Press.
- Walinsky, A. (1995, July). The crisis of public order. *Atlantic Monthly*, pp. 39–41, 44, 46–49, 52–54.
- Watson, B.C. (1992). The demographic revolution: Diversity in 21st century America. In B.J. Tidwell (Ed.), *The state of black America 1992*. New York: National Urban League.
- Webb, G. (1996, August 18, 19, 20). Dark alliance: The story behind the crack explosion. *San Jose Mercury News* (A special reprint), pp. 1–10.
- Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children—Revised*. New York: Psychological Corporation.
- Wechsler, D. (1981). *WAIS-R manual: Wechsler Adult Intelligence Scale—Revised*. New York: Psychological Corporation.
- Weizel, R. (1995, August 6). No boredom at this camp. *Boston Globe*, pp. 37, 40–41.
- West, C. (1994). *Race matters*. New York: Vintage Books.
- White, J.E. (1990, January 22). Genocide mumbo jumbo. *Time*, p. 20.
- Why more blacks do not donate organs. (1995, December 11). *Jet*, pp. 15–18.
- Wicker, T. (1996). *Tragic failure: Racial integration in America*. New York: William Morrow.
- Wilson, J.Q., & Herrnstein, R.J. (1985). *Crime and human nature*. New York: Simon & Schuster.
- Wilson, W.J. (1987). *The truly disadvantaged: The inner city, the underclass, and public policy*. Chicago: University of Chicago Press.

- Wilson, W.J., & Neckerman, K.M. (1986). Poverty and family: The widening gap between evidence and public policy issues. In S.H. Danziger & D.H. Weinberg (Eds.), *Fighting poverty: What works and what doesn't*. Cambridge, MA: Harvard University Press.
- Wolfgang, M.E., Figlio, R.M., & Sellin, T. (1972). *Delinquency in a birth cohort*. Chicago: University of Chicago Press.
- Yarrow, M.R., Waxler, C.Z., & Scott, P.M. (1971). Child effects on adult behavior. *Developmental Psychology*, 5, 300–311.
- Yussen, S.R., & Kane, P.T. (1985). Children's conception of intelligence. In S.R. Yussen (Ed.), *The growth of reflection in children*. New York: Academic.
- Zarefsky, D. (1986). *President Johnson's War on Poverty: Rhetoric and history*. University, AL: University of Alabama Press.
- Zobrist, A.L., & Carlson, F.R., Jr. (1973, June). An advice-taking chess computer. *Scientific American*, 228, 92–101.

## APPENDIX A

### CLARIFICATIONS CONCERNING THE POPULATION-IQ-OUTCOME MODEL

#### **Critical IQs and the Proxy Model**

The term *critical IQ* is used because it tests the model by representing a fictitious IQ threshold (hence, not a mean) that would account for a population's observed prevalence rate (viewed as a percentile) if everyone below that IQ met the criterion (i.e., became delinquent, contracted HIV) and no one with a higher IQ did. These conditions describe a step function, according to which the probability of qualifying goes from 1.0 for everyone below the critical IQ threshold to zero for everyone at or above it in a single jump or step. It is easy to see that when critical IQs for two populations with different IQ means and different prevalence rates match perfectly, as in Figure 1, leaving a difference of zero, the two step functions, which become vertical at precisely that IQ point, coincide.

Under the step-function interpretation of the model, the difference in prevalence rates for two populations whose critical IQs match can be understood, therefore, as being entirely attributable to the separation between their IQ distributions. The whole explanation of the difference in prevalence rates is simply that more of the population with the lower IQ mean extends below the critical IQ. As a description of real-life probabilities, the abrupt step-function interpretation is obviously totally implausible, but the job of models is usually to reveal some interesting aspect of reality more clearly, rather than to mimic all of its complexity, that is, to serve as a tool whose answers must be interpreted. In this case, the step function serves as a useful proxy model that enables us to test the responsiveness to population differences in IQ of a more realistic interpretation or probability model conveniently, if only indirectly. It accomplishes this by packing the total probability of the outcome in one tail of a population's IQ distribution, where

it can be compared easily with the similarly packed probability of another population, to see whether IQ commensurability holds.

### **The Realistic Model That Is Indirectly Tested by Critical IQs**

To illustrate the realistic form of the model, take delinquency as an example. I have argued on both theoretical and empirical grounds that the real probabilities within any single race would gradually decrease as IQ increases (recall the negative correlation at the individual level), but would also, however, have a downturn in the lowermost IQ range to reflect the special circumstances of retarded individuals, who may be either incapable of delinquency or not held culpable for misdeeds if capable of performing them (Gordon, 1976, 1986). In Washington, DC, for example, a mentally retarded person cannot be prosecuted or committed to a secure medical facility, even for self-confessed homicide (Castaneda, 1995). Herrnstein and Murray's (1994, p. 247) data depict the expected downturn in probability, beginning at IQ 75. A smooth curve fitted to and thus describing the real probabilities would constitute an empirical probability function different from the fictitious step function.

Empirical probability functions for other negative outcomes would normally be similarly gradualistic, except that the downturn in the lower IQ range need not be present. The probabilities that women age 19 through 23 have had an out-of-wedlock birth, for example, show a pronounced descent as  $g$  (measured by Armed Forces Qualification Test score) increases, but no downturn appears even in the lowest 5% of the test score distribution (Berlin & Sum, 1988, Figure 9). Both dropping out of school and receiving public assistance yield graphs similar to that for out-of-wedlock births (Berlin & Sum, 1988, Figures 6 and 7), and so does that for poverty among whites in 1989 (Herrnstein & Murray, 1994, p. 132). Extremely low IQ is evidently not a disqualification for experiencing these outcomes in the general population.

Probability functions like the one for out-of-wedlock births are analogous to the item-characteristic curves of psychometrics, which chart the probability of getting an item correct as a function of location on the latent trait, except that now the probability is of making what, for purposes of understanding, can usefully be viewed as a mistake.

Because IQ measurements on individuals are lacking, or because data are not reported in sufficient detail to draw race-specific smooth curves, or because samples are too small to provide stable probability estimates at each IQ interval, the realistic functions are typically unobserved. Unlike the step functions, exact locations of the realistic functions cannot be determined from prevalence rates alone, but it may be possible, as explained in the next section, to infer something useful about their relative locations for the two races. Ordinarily, the sign of the within-group correlations that such functions represent, that is, whether they are ascend-

ing or descending functions, can be inferred with confidence from IQ theory whenever the sign is not already known from individual-level data. But more can be inferred too.

### **Dependence of IQ Commensurability on Contextual Effects of IQ**

Recall that, for the proxy model, perfect matching of critical IQs means that the step functions for the two groups coincide exactly. Under such circumstances it is easy to apprehend visually that IQ is involved in some fundamental way for both races (see Figure 1). Differences in placement of the groups along the IQ continuum and the resulting differences in density of individuals located in higher risk segments of the IQ range would account totally for the difference in prevalence of whatever outcome is being examined.

The explanation of the model based on coinciding functions and differences in density cannot be the whole story for the realistic probability model, however. Perfect matching in critical IQs under the realistic model requires in addition that the two functions not coincide exactly. Instead, for negative outcomes such as those in Table 1, IQ commensurability in the realistic model requires the probability function for the lower scoring group to lie just the right distance, on average, above the function of the higher scoring group, neither too high nor too low.

The precise distance requirement stems from the fact that, when two critical IQs match and the two groups differ in mean IQ, it is mathematically impossible, according to simulations, for the two functions to coincide unless they are step functions, the limiting case (Gordon, 1987, pp. 35–37; see Hunter, Schmidt, & Rauschenberger, 1984, for a suggestive discussion of over- and under-prediction). This aspect of the model was anticipated but not fully recognized at first (Gordon, 1976, 1986), and hence too much emphasis was placed on the possibility of geometric congruence for the two realistic probability functions.

Simulations with plausibly shaped functions indicate that, when IQ commensurability holds, as much as 10% or more of the total Black prevalence rate could be attributable to separation between the Black and White functions, that is, to higher probability of crime or delinquency among Blacks than Whites for the same IQs. Because this separation between functions contributes to matching of critical IQs, it too can be understood as an IQ effect, one created by the differences between the two populations in the total IQ contexts of probands with the same IQ.

Just such a separation would appear between item-characteristic curves of unbiased intelligence test items if items were to be answered cooperatively by small groups within each race, because probands of the same IQ would have access to peers who were brighter on the average in one race than peers in the other race (recall the experiment by Laughlin and Johnson, 1966). Put simply, the average quality of within-race help would differ between races.

This basic fact concerning a race difference in quality of help is generally

acknowledged by social scientists, but without paying heed to population differences in intelligence as its fundamental source. Kahlenberg quoted two examples:

[According to] the National Research Council, "poor blacks, to a much greater degree than poor whites, interact mainly with other disadvantaged people. Black poor children attend schools with other poor children, go to churches with impoverished congregations, and deal with merchants geared to do business with a poor clientele." The University of Chicago's Mark Testa says: "Poor whites reside in areas which are ecologically and economically very different from poor [black areas]. Any observed relationships involving race would reflect, to some unknown degree, the relatively superior ecological niche many poor whites occupy with respect to jobs, marriage opportunities, and exposure to conventional role models." (Kahlenberg, 1996, p. 170, quoting from, respectively, Jaynes & Williams, 1989, pp. 283–284, and Wilson, 1987, pp. 58–60)

Although each taps important truth, both quotations illustrate the common practice of using wider contextual features of the sociological environment as an unacknowledged proxy, and therefore a circumlocution, for known differences not only in the intelligence context, but even for known differences in the intelligence of the probands within each context.

Maller's study of Health Areas in New York City, discussed earlier, provided a rare exception to sociological practice in that it rendered visible large differences in intelligence contexts among whites alone within a single urban setting. Recall, mean IQs ranged from 74 to 118. The delinquency contexts also varied enormously, by a factor of 11, from lowest nonzero rate to the highest. The correlation between mean IQ and delinquency rate at this level of neighborhood aggregation was  $-.57$  (Maller, 1933, Tables 3 and 4). Neighborhood IQ contexts nowadays would also differ among blacks too, as is implicit in the work of Wilson (1987), but they would also differ on average from those of whites.

Although they may sometimes prove difficult to separate in practice, the contextual and individual contributions of IQ can be kept analytically distinct. As the tentative 10% figure for delinquency suggests, the total contextual contribution can be modest in comparison with the individual proband's contribution, but the magnitude of the contextual component could vary widely, depending on the type of outcome.

When IQ commensurability holds for the proxy model, therefore, it holds in a more complex way for the realistic model by requiring a certain, nonnegligible separation between Black and White probability functions. Substantive interpretation of the separation between two realistic functions that exhibit IQ commensurability is crucial for understanding how such an effect could be produced. The problem is to explain why persons with the same IQ score can have systematically different probabilities of delinquency or other outcome in a manner that is entirely attributable to the IQs of the two populations. This problem is posed by the discovery of recurrent instances of IQ commensurability in prevalence data, and it is thus forced upon our attention rather than produced by pure theoretical speculation. Individual differences among probands alone cannot account for matching of critical IQs.

Whatever explanations for the separation between functions are advanced, they must account not just for the ordinal fact that one curve lies above the other, as introducing contextual variables might easily do (witness Kahlenberg's quotations), but for the fact that the contextual variables are evidently responsive to IQ parameters and so operate in a manner that maintains IQ commensurability (matches critical IQs) by achieving more or less the right distance between the two functions. This more stringent constraint distinguishes the contextual implications of the realistic model from all other contextual explanations, particularly the standard sociological ones, that predict only the ordering, not magnitude, of the difference: "Group A's average is lower than Group B's on X, therefore group A should obviously be higher (or lower) than Group B on Y."

### **The Components of Contextual and Population Effects**

Racial bias in law enforcement must be ruled out as an explanation of the precise distance requirement for delinquency, because any bias would have to be directed in some global way toward IQ, not race per se, in order for critical IQs to remain matched. If racial bias were present in its stereotypical form, there is no reason to expect that it would produce a fit to IQ differences. For these reasons, the first successful application of the model to delinquency data that preceded the Civil Rights era (for the years 1949–1954) led me to dismiss racial bias as an important factor in producing Black–White differences in rates of crime and delinquency (Gordon, 1976), although bias hypotheses were much in favor at the time. Since that time, research has pretty much eliminated net racial bias as an important source of racial disproportions in crime statistics (Tonry, 1995, chap. 2). The question of IQ bias in crime statistics will be revisited in Appendix C, after other possible causes of the separation in functions have been proposed.

Explanation of the vertical distance between the two race-specific functions in the realistic model amounts to a theoretical analysis or decomposition of the two race-specific probability ordinates for probands of any given IQ. Such ordinates represent the heights of the two functions at a given location on the horizontal IQ axis. Figure A-1 illustrates schematically the main details of such a hypothesized decomposition for a negative outcome such as criminality.

For the sake of visibility, the two ordinates representing the groups are placed side by side in Figure A-1, although for a given IQ, X, they would of course be superimposed. One ordinate represents individuals from a lower scoring population, the other represents individuals with the same tested IQ from a higher scoring population. The persuasiveness of the hypothesized decomposition depends on identifying IQ-related reasons, such as those to follow, for the difference in height between the two ordinates.

Reading the heights of the ordinates in Figure A-1 from the bottom up, the first component of probability derives from the individual's own observed IQ, just as one would expect. This component could be zero, but it would be the same for members of both populations. Probands with the same IQ experience the same

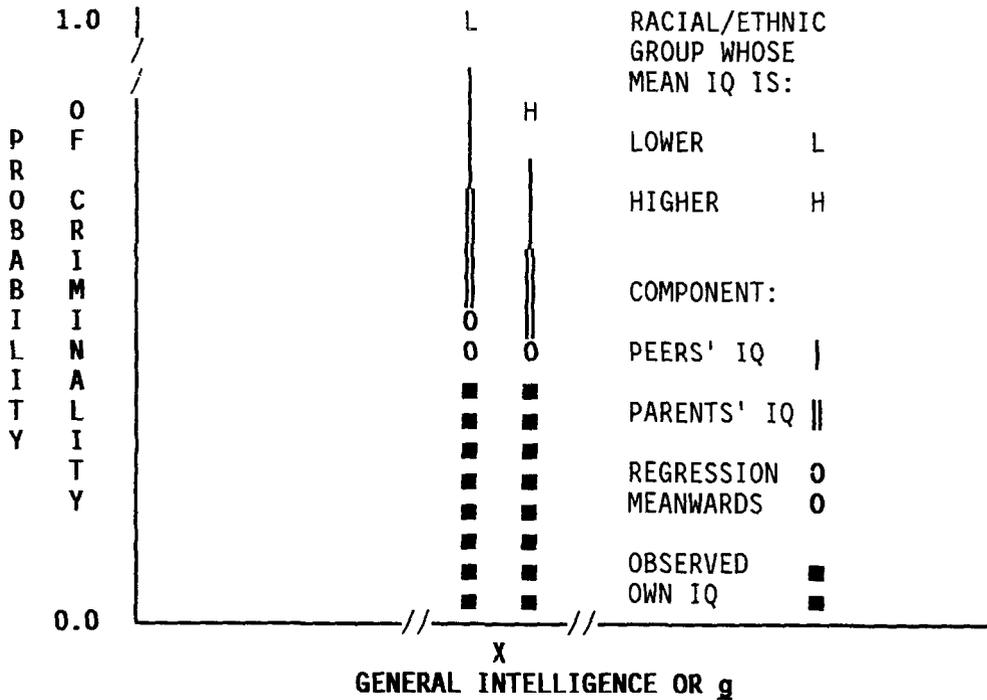


Figure A-1. Between-group differences in components of probability ordinate at IQ X.

increment in probability of criminality from this source alone, regardless of whether they are in the group with the lower or the higher average IQ. Clearly, this source cannot account for the mathematically required difference in height between the two functions and their ordinates at a given IQ point.

The second component reflects the fact that observed IQ scores are not true scores (not error free). Random measurement error in observed scores leads to some regression of true scores toward the mean. The true scores of individuals in each group regress toward their respective group means, producing differential regression even if test reliability is identical for both groups. Individuals in the lower group regress toward a lower mean, leading to a larger increment (or smaller decrement) in probability of the negative outcome for these individuals, because true scores, not observed scores, determine outcomes. The effect may be considered contextual or population level because it is population membership that establishes the mean to which scores of individual members regress, but the effect is not mediated through the *g* of other individuals, as with other typical contextual effects. If the individuals were members of some other population, they would regress to a different mean. (The magnitudes portrayed in Figure A-1 are intended only to make the different effects distinguishable from one another.)

Regression of true scores is thought to make a modest contribution to the explanation of the difference in height of the two ordinates, just as it does in linear

regressions for two groups with different means on predictor and outcome (Hunter et al., 1984). (Technically speaking, Figure A-1 portrays the net difference in probabilities attributable to regression for individuals at the same IQ score, because regression downward, for individuals located above the mean of their group, produces an increment in probability, as Figure A-1 seems to indicate, but regression upward, for individuals located below the mean of their group, produces a decrement in probability of the negative outcome.)

The third component derives from IQs of the parents (or other caretaking kin) of the probands and the role of parental intelligence in socialization and delinquency prevention. The strength of this component might vary, depending on whether adults or minors are under study. It is inferred from many studies of the family life and social background of delinquents that less intelligent parents are to some degree less successful at parenting on average and so contribute to a criminal outcome over and above the contribution of the offspring's IQ (Gordon, 1976). Such an inference hardly seems problematic in view of the correlation of IQ with SES and of SES with delinquency (Gordon, 1967, 1986), not to mention the genetic linkage between parents and their delinquent offspring, who tend to have below-average IQs. Parenting is, after all, a kind of service job, and *g* is the best single predictor of performance in a variety of jobs (Gottfredson, 1986b).

Typical findings are that discipline, for example, in the families of delinquents is very often lax, harsh, or erratic, and supervision unsuitable (Glueck & Glueck, 1950, Tables X-10, XI-22). Burt (1925/1948, p. 96) stated that of all environmental conditions in his list of causes, "the group showing the closest connexion with crime consists of . . . defective discipline." Burt named overstrictness, laxness, and virtual absence of home discipline. Certainly, evidence for family process variables as a contributing factor to delinquency is not lacking (Blumstein et al., 1986, pp. 43-44; Sampson & Laub, 1993, chap. 4).

Evidence that mothers with lower measured IQ make less effective parents has been reported by Herrnstein and Murray (1994, chap. 10), but not all of their outcomes for children are reasonably free of effects from the child's IQ, which is correlated with mother's IQ. The effects from parent IQ and offspring IQ on successful socialization may be correlated and reciprocal, and thus inextricable from each other without very close observation indeed (Bell, 1968; Yarrow, Waxler, & Scott, 1971). For that reason, I emphasized the IQs of the parent-child dyad as the fundamental unit of observation for this aspect of the model (Gordon, 1976, pp. 267-269). Nontrivial intergenerational regression toward the mean is to be expected in observed IQ unless parent and offspring IQs are perfectly or very highly correlated, which they are not. (Regression works in either direction, child to parent or parent to child, depending on which is in the role of proband.)

Figure A-1 reflects the fact that the parental regression too is differential, again because population means differ by race. For example, Black probands of IQ 90 will on average have parents whose midparent IQ is lower than theirs, and White probands of IQ 90 will on average have parents whose midparent IQ is higher than

theirs, as an IQ of 90 falls above the Black population mean and below the White population mean (for means, see Table 2). Although the direction of regression meanward may thus differ at various points in the IQ range, the magnitude of the expected race difference in parental IQ would be constant throughout the range of offspring IQs. The observed scores of parents (or children) would also be subject to regression toward their own true scores.

According to one set of reasonable assumptions, the systematic difference in midparent IQ for Blacks and Whites that is attributable to regression could be about 5.3 IQ points (Gordon, 1976). This amounts to 84% of the average class interval when occupations are stratified in five categories and mean IQs of their incumbents compared (Reynolds et al., 1987, Table 4, ages 20–54). As Figure A-1 indicates, the net difference in probability of a negative outcome from this source is consistently less favorable to the racial group with the lower IQ mean.

The fourth component concerns peers, the remaining major category of actors in the social environment whose IQs could prove relevant (Sampson & Laub, 1993, chap. 5). IQs of peers are subject to regression effects in both obtained and true scores too, which, for simplicity, are not separately distinguished in Figure A-1. Finding peers of similar or lower IQ may be easier in one distribution than in the other simply because the probability density of those IQs is greater in one race than in the other. The crucial point, of course, is not that the peers have an average IQ lower than the proband's IQ (they may not), but that they have an expected average IQ lower than that of peers of a proband of the same IQ in the race with the higher mean. Because of the difference, and the fact that mixed-race offending is infrequent (Reiss, 1986, p. 133), lawbreaking would be facilitated more easily in one population than the other. Victimization data bear this out by indicating that Blacks (60%) are less likely than Whites (72%) to be solo offenders when the crime is one of violence (Reiss, 1986, p. 133), perhaps the hardest kind of crime for which to recruit, and, as is known, rates of violent crime are much higher among Blacks.

Certainly, the availability of delinquents and delinquents-to-be with whom to team up is three or four times greater among Blacks than Whites, as Table 1 indicates. According to one set of data, 73.3% of robbery offenders had one or more accomplices (Reiss, 1986, Table 3), only a fraction of whom could possibly be instigators (simultaneous mutual instigation presumably being a rarity). Sometimes accomplices barely know one another, but often are acquired easily enough through loose networks in lower class communities for temporary relationships in offending (Reiss, 1986). Readers are invited to consider how easy or difficult it would be to recruit an accomplice from among their own peers to help commit the next serious multiple-offender crime they see reported.

### **The Role of Race and Generations in the Model**

Combined, the various components identified in Figure A-1 indicate a systematically higher probability of the negative outcome for all members of the lower group than for their IQ-matched counterparts in the higher group. This, of course,

is a key point for the realistic model. The total vertical difference in probability is systematically related to the horizontal difference between races in IQ parameters because the same IQ *means* (to within an additive constant) and *SDs* describe the distributions of parents and of peers as describe the distributions of the probands. The three IQ distributions within race are linked, therefore, no matter what the values of the correlations between IQs of individuals in one distribution (say, probands) and IQs of individuals in another distribution (say, effective peers) happen to be. How necessary it is that these correlations be roughly the same in both races in order for the model to work remains an unresolved question. Correlation matrices are generally fairly similar across race.

To repeat, locations in all three distributions (proband, parent, peer) are probabilistically linked to one another within race by essentially the same IQ parameters, which would account for observed IQ commensurability under the assumption of separated probability functions in the more realistic model. To the extent that race represents a somewhat closed system of kinship and important nonkin relationships, the model probably functions better. Because of the linkages between IQ locations in the three identified distributions, the model is in varying degrees a two-generation one at least and so may apply better to race than to other forms of grouping, which might not assure, for example, that but one set of IQ parameters applies to both proband and parent generations. Social class groupings within race would not possess the same IQ parameters for parents and offspring, because of intergenerational regression meanward of uncertain and perhaps varying magnitude, depending on the childhood origins of the parents (how vertically mobile they were). No expectations seem warranted, at this point, that the model will hold for other than relatively discrete and well-characterized racial or ethnic populations.

### **A Helpful Between-Group Analogy and Misconceptions About the Model**

A simple analogy may help to explain the greater prominence that successful applications of the race–IQ–outcome model assign to *g* in the case of between-group differences, in contrast to within-group models. Height may be an important predictor of scoring in professional basketball, but height must compete with other important athletic variables to explain within-team scoring differences. Some talented short players, like the 5'3" Muggsy Bogues, for example, can jump unusually high for their height and thus contribute to their team's scoring beyond what their height predicts. Indeed, when used in selection, such offsetting considerations such as jumping ability and height can actually reduce the individual correlations between all predictors and a complex outcome, because outstanding scores on any one can lead to the selection of valuable players with less impressive scores on other predictors. Constitute two teams so that the only systematic difference between them is one of height, however, and height will emerge as the total explanation of their scoring differences.

Conversely, when height does emerge as the total explanation by predicting score differences in advance, it suggests without further ado that there are no other relevant between-team differences that are independent of the height parameter for teams. A putative competing explanation of scoring differences, such as an average difference in rebounding proficiency, may now be only another manifestation in different form of the height difference in the specified between-groups situation (but certainly not so in the within-groups context). In modeling jargon, a between-team difference in rebounding success may, under the circumstances, be merely an effect that is co-endogenous with the between-team scoring difference, with both differences the result of height, rather than a contributing explanation in its own right to that scoring difference.

Least-squares models that address within-group and between-group variance simultaneously, and with multiple independent variables, in order to predict individual level differences may actually be misspecified for the purpose of understanding between-group differences, because they dilute the apparent importance of what could prove the key explanation of many between-group differences. An explanation that accounts for all of the between-group variance may account for only a small part of the within-group variance and, hence, only a fraction, too, of the total variance. When that explanation involves differences in intelligence, we have one more example of an impediment to the full appreciation of the role of intelligence in everyday life. Manifestations of group differences in  $g$  can be very much a part of life, but unrecognized as such because they appear under numerous guises.

There is a totally mistaken view, based on a misconceived application of the usual regression model, that the within-group, individual-level, correlation between IQ and the outcome must be perfect,  $r \approx \pm 1.0$ , and thus account for 100% of the variance if the population-IQ-outcome model (i.e., IQ alone) is to account totally for the difference in the outcome variable between groups. Because perfect within-group, individual-level, correlations are implausible as well as often inconsistent with empirical estimates of their magnitude, this argument was employed to dismiss the population-IQ-outcome model of between-group effects out of hand.

It is easy to verify, however, that even the step function, proxy version of the model, as visible in Figure 1, not only does not require perfect individual-level correlations between IQ and outcome, but actually entails point biserial correlations (i.e., at the individual level) that are often much smaller than  $\pm 1.0$ . It is possible to calculate such point biserial correlations under the step-function model for any prevalence rate (the means of sections of the normal distribution required for the calculation can be obtained as described in Kelley, 1947, pp. 295–297). For prevalence rates of 1%, 4%, 17%, and 50%, for example, the corresponding individual level correlations would be  $-.32$ ,  $-.44$ ,  $-.67$ , and  $-.80$  between IQ and outcome. These correlations are well below  $-1.0$  in absolute size and entirely compatible with IQ commensurability, as prevalences with fairly similar magnitudes in Table 1 testify in Table 2. The mistaken demand for perfect correlation

represents, in part, adherence to an individual-level model of the usual sort, which does not admit of different processes or effects at the between-group level.

Point biserial correlations between IQ and outcome can be viewed as proxies for  $g$  loadings of the outcome. This relation makes clear that degree of fit need not vary with the  $g$  loading of outcomes (items), as fit depends only on the prevalence rates, which correspond to the relative difficulty levels of outcomes (now viewed as though they were items) in the two populations being compared. If IQ, like the outcome, is dichotomized too, a prevalence and its complement become simply one set of marginal percentages in the resulting fourfold table, for which the correlation is a phi coefficient. Because phi correlations are largely independent of both marginals in fourfold tables, various sets of outcome marginals that figure in perfect fits of the model are thus compatible with a wide range of within-group phi correlations between IQ and outcome. In other words, there is no necessary relation between the fit of prevalences in the model and the  $g$  loadings of the outcome. Between-group disturbances from other variables and sampling errors are far more likely sources of variations in fit than such  $g$  loadings.

Technically, the two within-group correlations, phi or point biserial, accompanying any perfectly fitting pair of race-specific prevalences could be zero or even opposite in sign to the intuitively likely sign under a parsimonious theory. Such hypotheses would seldom make theoretical sense, however, and would likely be refuted by actual empirical evidence, as in the case of delinquency and other outcomes already mentioned, where the signs of the within- and between-group correlations were consistent with each other. A possible example of perversely signed within-group correlations is discussed in connection with Table 3 (see Appendix D), which deals with opinions rather than performance behaviors. Opinions, because of their malleability by forces other than reason and practicality, even among the very bright, may be the one domain in which such counterintuitive examples ever appear.

## **APPENDIX B**

### **AFFINITY OF THE POPULATION-IQ-OUTCOME MODEL TO EPIDEMIOLOGICAL MODELS**

Social network modelers may recognize affinities between the account of the Black and White probability functions in the realistic model provided in Appendix A and a wide class of well-studied models regularly employed in epidemiological research. A brief sketch of these affinities may help make the model and its mechanics more understandable to others, and perhaps may interest mathematical sophisticates in developing models of the population effects of  $g$  on a wide range of outcomes. Much of what is covered in this appendix will serve to unify the second level of analysis of IQ effects, namely, the level of the local interpersonal context of individuals, with the third or population level.

The following account draws heavily on an article by Kretzschmar et al. (1994) for details common to epidemiological network models, which classically assume that a population contains two classes of individuals, infectives ( $I$ ) and susceptibles ( $S$ ).

### **Contagion**

Epidemiological models further assume “that the probability for an infectious contact [now, outcome-promoting contact] is proportional to the number of infectives [now, risk-enhancing individuals] in the population” (Kretzschmar et al., 1994, p. 564). The per capita rate of infection (now, rate of outcome production) is determined by multiplying the number or density of infectives,  $I$ , by a constant,  $\lambda$ , which combines the rate of contact and the probability of transmitting the disease (enhancing the risk) per contact. In the population–IQ–outcome model, however, the transmission probability of a negative outcome is, in addition, a continuous descending function of the IQ of the potential risk-enhancer (which is not to deny that influence on others, given that an individual does enhance risks, may increase with that person’s IQ). Another important distinction, not appropriate for infection models, is that the IQ model allows for a certain probability of “spontaneous generation” of the outcome in individuals as probands, influenced by their own IQs (and also other variables external to the model).

Epidemiologists regard the number of contacts per unit time as being proportional to the number or density of susceptibles, “so that the transition from class  $S$  to class  $I$  [say, from not delinquent to delinquent] is described by  $[\lambda I(S)]$ ” (Kretzschmar et al., 1994, p. 565). Susceptibility in the population–IQ–outcome model would, in addition, be a descending function of IQ, although in the case of a true infection, such as HIV, the usual biological considerations would also apply.

### **SIR Model**

The epidemiological model is known as the SIR model, because it adds to the  $S$  and  $I$  classes an  $R$  group that stands for the number or density of infected individuals removed by death or immunity following recovery (e.g., execution or imprisonment for lawbreaking or, to take a different outcome, for holding a forbidden opinion in a totalitarian society). Although the  $R$  concept often may not apply outside of a disease context, it can be illuminating to consider that in some cases it might apply.

### **The $R$ Effect in Criminality**

A huge reduction in victimization rates determined from annual National Crime Surveys of households between 1973 and 1989 coincided with markedly increased rates of admission to prison of persons arrested (Langan, 1991). There seems to be little hard evidence for deterrent effects (Brier & Fienberg, 1980; Loftin, 1980), which may not be all that surprising in view of the dependence of deterrence on both look ahead (“How will I feel in my  $n$ th year of imprisonment?”) and making good use of information (recall the high  $g$  loading of the

Wechsler Information substest). Nevertheless, incapacitation and deterrence were thought to be heavily responsible for the 1973–1989 drop in victimization rate (Langan, 1991). The *R* concept and the realistic version of the population–IQ–outcome model alert us to the possibility that the removal of criminals from the general population (incapacitation) also removes their contributions as peers to the probability of criminality by susceptibles still at large.

An *R* effect in a natural experiment that would help confirm the peer effect is by no means out of the question. In Chicago, years ago, social workers assigned to gangs would secretly “cut loose,” as they euphemistically put it, a member of the gang who persisted in committing strong-arm robberies by informing the police, who could then place that individual in a line-up for possible identification by victims. The aim of the gang workers was to keep the serious offender from involving other gang members in similar crimes.

Young high-rate offenders tend to have the most accomplices, leading one criminologist to conclude:

Were one able to select these high-rate offenders prospectively and isolate them, one could, in the short run, avert a substantial amount of juvenile crime, both those they commit alone and those attributable to recruiting others. Moreover, to the degree that their incapacitation deters or *reduces* [italics added] the offending rate of a sizable proportion of their accomplices, considerable additional reductions in crime might be expected since it seems likely these high rate offenders seek accomplices who might otherwise not be as active in offending. (Reiss, 1986, p. 144)

Desistance of offending can sometimes be attributed to the disruption of affiliations with other offenders, perhaps through the replacement of those affiliations by ones with law-abiding persons, such as wives, employers, and educational institutions (Reiss, 1986, pp. 149–152; Sampson & Laub, 1993, p. 203).

### **SIR Model and Law of Mass Action**

A fundamental assumption of the SIR model is that the transition from class *S* to class *I* described by  $\lambda(I)S$  “is analogous to what is known in reaction kinetics as the law of mass action: the reaction velocity is proportional to the concentrations of the two reacting substances” (Kretzschmar et al., 1994, p. 565). Homogeneous mixing of the particles of the two substances is assumed. “The same applies to the epidemic model: the SIR model assumes that the probability of contact between any two members of the population is the same at all times [i.e., for all pairs]” (Kretzschmar et al., 1994, p. 565). The assumption of homogeneous mixing may apply well to the spread of some diseases under certain conditions, such as measles within a school classroom, but not to the spread of HIV, “where selective mixing within and among groups in the population plays a predominant role” (Kretzschmar et al., 1994, p. 565).

The problem of selective mixing is important for the population–IQ–outcome model too, because social life is known to be characterized by selective and often correlated mixing according to IQ. Such correlations, however, could serve as a basis for mathematical simplification by assigning predicted average values, say, of peer and parent IQ to probands using the probands' IQs as a predictor, thus summarizing the IQ values of many contacts. It would remain, then, simply to estimate the average effects associated with such average IQ values of persons in the probands' settings. In any case, when the model fits, that can be taken as a sign that it is robust to departures from random mixing.

If necessary, the law of mass action can be modified and the assumption of equally probable mixing relaxed by, for example, compartmentizing a population into subgroups and considering separately the probabilities of contact within and among the subgroups, perhaps using a procedure involving "mixing matrices" (Kretzschmar et al., 1994, p. 565; Schmitz & Castillo-Chavez, 1994). Although the question of how finely IQ subgroupings could be divided without the model becoming mathematically intractable would remain to be determined, the general idea behind such a solution makes it easier to understand how the population–IQ–outcome model might work in detail in the real world, even if full subdivision were not carried through in the model. A requirement would be that the subgroups sum to the entire population, so that population IQ parameters continue to apply.

The SIR model is the basis for social contagion models of first sexual intercourse explored by Rowe and his associates (Rowe & Rodgers, 1994; Rowe, Rodgers, & Meseck-Bushey, 1989), who found their models of that outcome to be robust to departures from the assumption of random mixing. Their work, discussed in Appendix C in another connection, demonstrates that the SIR model can be applied to outcomes other than disease.

## APPENDIX C FURTHER DETAILS OF DELINQUENCY AND CRIME RELEVANT TO THE IQ MODEL

### **Resolution of Anomalous Results Leads to the Deviance Principle**

Establishing the comparability of the synthetic cohort and birth cohort methods for determining prevalence helps to resolve a longstanding anomaly concerning use of the population–IQ–outcome model. Race-specific male prevalences based on police contact records for birth Cohort I in Table 1, first reported by Wolfgang, Figlio, and Sellin (1972, p. 54), were found to be anomalous when compared with juvenile court record prevalences in line 1 of Table 1 for the same city derived using the synthetic cohort method (Gordon, 1976, pp. 215, 237–240, 271). Wolfgang et al.'s police contact prevalences for Blacks (50.24%) and Whites (28.64%) by age 18 failed to fit. There were either far too many Whites or far too few Blacks qualifying as delinquent for the model to work well. One could not be sure

whether this was because of the difference in methods for determining prevalence, the difference in criteria of delinquency, or some failure of the model.

Tracy, Wolfgang, and Figlio's (1985) further analysis of the Cohort I data and their addition of the birth Cohort II data have made it possible to untangle the anomaly concerning fit, because now an excellent fit can be observed in Table 2 when the broad criterion of police contact for Cohort I is narrowed to Uniform Crime Report (UCR) index offenses only. The UCR category contains only the more serious infractions (listed in Table 1). Tracy et al.'s Cohort I and Cohort II prevalences for milder nonindex offenses (not presented here) continue not to fit the model, just as did Wolfgang et al.'s (1972) police contacts for all nontraffic offenses (see Visher & Roth, 1986, Table A-1) and Hirschi's (1969, Table 14A) police contact data for juveniles in California, all of which had similarly seemed anomalous. The various results indicate that more severe and hence more deviant forms of criminality are what the model fits best. The anomaly was thus caused by the difference in severity of criteria.

In a different domain, a similar effect of outcome severity on the model has been detected in data analyzed for other purposes by Rowe and his associates (Rowe & Rodgers, 1994; Rowe, Rodgers, & Meseck-Bushey, 1989). Their two attempts to model Black-White differences in the age-specific onset of sexual intercourse among adolescents did not consider intelligence, but relied instead on assumptions about race differences in age of female sexual maturation and in rate of contact. Their model met with qualified success when applied to data from two separate studies, but application of the population-IQ-outcome model to Rowe et al.'s prevalence data also achieved a qualified success, and the nature of that success holds a further clue as to the kinds of behavior for which the model is appropriate.

Briefly, the IQ model worked well at youngest ages 12 and 13 for males and females in the early study by Rowe et al. (1989, Table 1), and then failed progressively worse at later ages to 16 for both sexes. It also worked well for males at the youngest four ages (10-13) in a later study by Rowe and Rodgers (1994, Table 4) and again failed progressively worse for both sexes to age 18. The vital clue here is that loss of virginity becomes less and less deviant or socially unacceptable, less and less a mistake, so to speak, with age. Eventually, remaining a virgin could be considered unusual, although not necessarily a mistake. Evidently, the degree of mistake represented by a deviant outcome can moderate the applicability of the population-IQ-outcome model. The generalization emerging from such observations, identified as the deviance principle, is considered further at a later point in the main text.

### **Commonalities Among Early Intercourse, Delinquency, and Other Forms of Deviance**

Early onset of sexual intercourse has been found repeatedly to be associated with, and even predicted by, problem, antisocial, and delinquent behavior in adoles-

cents of both sexes, even when the ages researched do not extend as low as 12 to 14 (e.g., Capaldi, Crosby, & Stoolmiller, 1996; East, 1996; Tubman, Windle, & Windle, 1996). The fit of the population–IQ–outcome model to the more deviant forms of both kinds of behavior, therefore, can be placed within a more-inclusive framework. Juveniles at risk for either mistake tend also to be at risk for the other. A similar broad association holds between drivers with records of repeated accidents, on the one hand, and contacts, on the other hand, with various agencies that try to remedy other kinds of mistakes, such as juvenile and adult courts, social service agencies, credit and collection agencies, and venereal disease clinics (McFarland & Moore, 1961, Figure 3). Recall, judgment had been implicated in vehicle accidents by Iskrant and Joliet (1968; O’Toole, 1990), and the need for judgment is known to be general.

### **The Differential Detection Hypothesis Revisited**

The racial pattern described here for mild criteria of offending, for which Whites are overrepresented or Blacks underrepresented according to the model, is oddly inconsistent in direction with what criminologists call the *differential detection hypothesis*: namely, that low-IQ offenders get caught more easily and hence that is what accounts for associations between crime and IQ, and perhaps also, in view of the IQ difference, for racial disparities in official crime statistics. Police contact and arrest are precisely the points at which detection comes most purely into play, but it is generally the view now that “arrests can by and large be taken as reasonable reflections of the involvement in serious crime of members of different racial groups” (Tonry, 1995, p. 71). According to the differential detection hypothesis, and given the IQ means of both populations, Blacks should not be underrepresented at the point of police contact or of arrest nor Whites overrepresented, relative to IQ, if IQ affects detection. Basing the arrest on UCR index offenses, however, which are more severe, led to a perfect fit (within rounding error) for the very same individuals (line 9 of Table 2 vs. line 8).

The two anecdotes concerning incompetent robbers might seem to lend credence to the hypothesis that IQ fits prevalence data based on official justice system statistics only because smarter offenders are less likely to be detected. However, relevant research findings, in addition to the kind mentioned, are inconsistent with this hypothesis.

First, differences in getting caught on the basis of intelligence level would seem to figure most in the earliest stage of official processing, yet the model fits later stages too (and often better, as we have just seen), such as incarceration (Table 2), despite the fact that the prevalences (and thus the percentages of base populations referred for further processing) drop sharply at each stage. Although the mean IQ of offenders does decrease at each later stage of official processing (Herrnstein & Murray, 1994, p. 246), one would have to assume that individuals were selected for further processing on the basis of IQ itself to account for the continued good fit, yet there is no basis for such an assumption. The decrease of

IQ over the successive steps of processing is likely related to seriousness of present offense and of past record, which can warrant the additional processing and also signal lower IQ.

Second, negative IQ–delinquency, IQ–criminality correlations appear also in self-report data on offending (Gordon, 1986, pp. 116–117; Herrnstein & Murray, 1994, pp. 247, 249; Hirschi & Hindelang, 1977; Moffitt & Silva, 1988). Self-reports do not depend on official processing or detection.

Third, when investigated, no IQ difference between caught and uncaught individuals who were equally delinquent according to self-report and other ratings was found (Moffitt & Silva, 1988). Both caught and uncaught delinquents had lower IQ averages than nondelinquents.

Finally, determinations of race of offender from national surveys of victims of rape, robbery, and assault reveal Black/White offender ratios that are consistent with the large ratios observed in official data and on which the success of the IQ model in Table 2 depends, although, again, no official processing had been involved in producing the eyewitness data (Gordon, 1986, pp. 100–101; Laub, 1983).

#### **Additional Comments Concerning the Prevalence of Lawbreakers**

In criminology, a coefficient  $\lambda$  is used to express individual differences in frequency of offending, which varies greatly from criminal to criminal. Criminology's  $\lambda$  should not be confused with the  $\lambda$  of epidemiology, discussed in Appendix B, although the two could easily prove to be empirically related. The former is a rate of criminal activity; the latter embodies a rate of contact, not with victims, but with susceptibles to offending or to whatever the outcome happens to be. It is relevant to note that Black/White ratios for the  $\lambda$  of criminology are only about 1:1 to 1.3:1 for most crimes and are virtually identical for the more serious UCR index crimes (Cohen, 1986, Table B-31). This means that the large race difference in aggregate crime and arrest is mostly attributable to a Black–White difference in what criminologists (Visher & Roth, 1986) now prefer to call *participation*, which corresponds exactly to prevalence as used here, rather than to a difference in rate of offending for offenders that causes offenders in one race to be apprehended more often than offenders in the other. It is not the case that Black offenders are much busier than White offenders; rather, there are more Black than White offenders per capita. To avoid the confusion of using two terms for outcomes (participation for crime, prevalence for noncrime) and to maintain focus on the applicability of the population–IQ–outcome model to a variety of outcomes and to well-defined populations, I use the term *prevalence* throughout.

Participation serves as a more convenient term in special contexts, such as participation in crime by members of an ad hoc sample or within a family, where applying the concept of prevalence might seem overreaching. Thus, all participation estimates are not necessarily estimates for well-defined populations (e.g., Visher & Roth, 1986, Tables A-1 & A-2).

The small difference between Blacks and Whites in criminology's lambda, as already noted and further similarity between Blacks and Whites in arrest rates defined as arrests per year per active offender, a crime statistic called mu, as well as in degree of specialization by type of offense (Blumstein et al., 1986, pp. 4-5 & Tables 3-9 and 3-12; Cohen, 1986, Table B-19), help to explain the good performance of prevalences for lawbreakers as input for the population-IQ-outcome model. Arrests per year per offender have rates high enough to ensure a high probability that most offenders are apprehended eventually during even a relatively brief criminal career (like the probability of losing at Russian roulette if played repeatedly), and so when the period of risk is one of years, lifetime prevalences defined by official records tend to smooth out any remaining perturbations from such minor differences as may remain.

#### **APPENDIX D**

### **METHODOLOGICAL ISSUES PERTAINING TO PUBLIC OPINION DATA**

#### **Right Answer Analysis and Other Procedural Decisions Required**

The multicategoric formats supplied for responses to opinion items require some decisions before analysis can proceed. First, it must be decided which is the more plausibly incorrect response. For beliefs such as those in Table 3, this is not a difficult decision, even though it may remain a tentative one in view of the uncertainty surrounding claims of conspiracy. Presented with the poll results, Harvard psychiatrist Alvin Poussaint, himself a Black, seemed to concur. He expressed dismay especially over Rumor 3, concerning AIDS, "What do they think? That there's a Nazi regime?" (DeParle, 1990, p. B6). In any case, if one makes the wrong decision for the data in their particular time and place, the model simply will not work, as it will entail what amounts to counting percentiles from the wrong end of the IQ distribution when calculating critical IQs and hence produce a disastrous fit. This is not to claim, however, that an answer that works well in the model establishes the truth of the matter in question, although it may add to the a priori probability that truth lies in the direction so indicated.

Second, it must be decided how to deal with "don't know" (DK) and "no answer" (NA) categories that are present for all items. My position is that they must be included so as to account for the full distribution of a population.

Third, it must be decided whether DK, NA, and sometimes "Not enough information to say" are to be classified as wrong or right responses. This depends on the question. My view is that such responses to questions like those in Table 3 are good candidates for wrong answers, just as they would clearly be, for example, to a question beginning, "Some people say the moon is made of green cheese . . ." True, one must have some general savviness about how the world works and about the vulnerabilities to unravelling of complex conspiracies to discount these beliefs, but that only helps to explain why the IQ model might fit. Savviness of

exactly that kind was displayed in a *New York Times* editorial, for instance, on similarly conspiratorial themes presented as history in recent Oliver Stone films. Experience, the *Times* argued, showed that governmental agencies lacked the competence and managerial expertise to carry out such elaborate and highly charged plots, even if they wished to do so ("Oliver's World," 1995). A critique of genocidal conspiracy rumors relying on more ordinary forms of common sense, which also supports the right answer analysis already offered, appeared in *Time*, entitled, "Genocide Mumbo Jumbo" (White, 1990, p. 20; see also the interview with White in Allen, 1990). That article argued, for example, that Black mayors and chiefs of police would certainly be in a position to expose any government conspiracy to distribute drugs in Black neighborhoods. Later, the prosecution in the O.J. Simpson trial would invoke the same logic in response to defense allegations that the police were both incompetent and engaged in a conspiracy to frame Simpson: Attorney Marcia Clark argued that the defense "jumped from we are stupid bumlbers [to] we are brilliant conspirators" ("Excerpts," 1995, p. 8), or, as a *New York Times* editorial put it, "the defense claims of incompetence and conspiracy were at war with each other" ("Race Cards," 1995, p. 18; see also Toobin, 1996, p. 335).

Fourth, multicategoric or polytomous main responses must be reduced to dichotomies in order to apply the IQ model, and this requires deciding where to place the cutting point. Survey organizations sometimes package all *agree* or *disagree* responses, when there are degrees of each, under the headings "net agree" and "net disagree." This seems more reasonable than cutting at extreme *agree* or extreme *disagree* categories, although one is free to explore that option.

Right answer analysis of rumors depends more on common sense than on strict certitude. For example, during August 1996, a sensational news report claimed to link the U.S. Central Intelligence Agency (CIA) to the inception of cocaine peddling in Black sectors of Los Angeles and to the consequent spread of the crack epidemic to Black communities in the rest of the nation during the 1980s (Webb, 1996). The allegation was taken up, with unseemly haste (Raspberry, 1996), by Black leaders, journalists, and talk-show hosts as the explanation for the disastrous impact of drugs and crime on Black communities everywhere (Fletcher, 1996). As widely purveyed, the story appears to pose a challenge to the assumed incorrectness of Rumor 2 in Table 3.

Examination of the source reveals that the story is crafted out of innuendo and implied guilt by association, and it contains no actual claim of CIA involvement or of intent to harm Blacks (the alleged motive was to raise money for the Nicaraguan Contras). As one critical article pointed out, "The series doesn't actually say the CIA knew about the drug trafficking [by convicted dealers with Contra connections]," and its author conceded as much when interviewed (Kurtz, 1996, p. B1). Major follow-up articles in the *Washington Post*, *New York Times*, and *Los Angeles Times* noted numerous flaws in both the narrower and the wider allegations, and reported that the author of the series had promoted the idea of a CIA

connection to the defense lawyer of a defendant on trial for drug dealing prior to publication of the series and then used testimony from the trial to bolster his published story (Fletcher, 1996; Golden, 1996a, 1996b; Katz, 1996; Suro & Pincus, 1996). Several government investigations are under way, and it remains to be seen whether anything is uncovered that lends substance, as distinct from credence, to Rumor 2 in its stated form, and whether, if there is any substance, the CIA was centrally involved or just some peripheral agents with bad judgment. Given the clandestine nature of CIA work, it will not be easy to put the story to rest in any case.

### **Use of Polling Data Entails Reliance on Much Smaller Samples, With New Sources of Error**

Application of the IQ model to opinion data requires that certain features of survey research be noted. Far smaller samples are now involved than was the case for the prevalence data in Table 2. Margins of error, usually defined as the 95% confidence interval determined for a response of 50%, indicate the probable ranges of sampling error. Telephone polling may be subject to biases for sampling persons representative of the IQ distribution, as they tend to have response rates about 5% lower nowadays than in-person interviews and to lead to some undercoverage of the household population (e.g., "Telephone Helps," 1978). Polling organizations caution, too, that sources of error other than sampling can affect poll results, such as interviewer effects; the wording and sequence of questions; the sequence of options in a given question (Moore, 1995); public events occurring during fieldwork; time allowed for fieldwork, such as only one night, thus limiting callbacks, which may vary from zero to six or more (Taylor, 1995); differences between polling organizations in callback policies, about which they tend to be tightlipped; reliance on less-preferred cluster sampling; and undetected flaws in procedure. Differences from survey to survey in items that precede a given item that is repeated can thus influence responses to the repeated item. The use of random sampling techniques is, of course, one of the great strengths of survey research, but that strength should not blind us to new sources of potential error.

### **Understanding the Possibility of Opposite Signs for Within-Group and Between-Group Relations With IQ**

For purposes of her own, Hochschild (1995) obtained special analyses of the rumor data in Table 3 that are interesting to consider from the standpoint of IQ, even though her results were cast as education or class differences, because they seem to suggest an IQ correlation within groups that would be opposite in sign to the IQ correlation between groups. Such a seemingly contradictory pattern of correlations is mathematically possible, as was noted in Appendix A, even when the model fits well.

Without holding age constant, in view of steady increases in amount of schooling throughout this century, relying on years of schooling as a proxy for IQ is

risky, because age-cohort effects in education could be mistaken for and thus dilute IQ effects (e.g., Reynolds et al., 1987, Table 8). For the sake of discussion, however, let us assume that Hochschild's results do reflect positive correlations between IQ and acceptance of the three rumors within each race, despite reverse correlations with IQ between race.

Hochschild (1995, Table 5.1) compared respondents in each race who had less than a high school education with those she labeled simply "college." Omitted were high school graduates, whose unknown responses might qualify the picture by falling in a different order. In any case, Hochschild found that Blacks in the college category accepted Rumors 1–3 in Table 3 more often than Blacks with less than high school education, and the same held true for Whites and Rumor 1, with Rumors 2 and 3 only narrowly pointed in the other direction. Within her categories of education, of course, Blacks accepted the rumors far more often than Whites, which is consistent with the between-group IQ model, even though the within-group correlations implied by the percentages do not always bear a sign that would be similarly consistent.

For comprehending Hochschild's within-group results, it is useful to recall George Orwell's (1946/1968, p. 173) comment, when referring to both polls and IQ, that there are certain questions to which "the less-gifted person would have been likelier to give a right answer." Distorting currents of complex thought or subcultural ideology to which intellectuals, a group observed keenly by Orwell throughout his career, are more exposed than working-class persons could account for the anomalous effect. The uncritical oppression–victimization theory presented in much college teaching about race issues would certainly represent such a distorting current of thought for many in the college group. When mistaken, prevailing opinion among influential elites can add a component of difficulty to any question that might well be attenuated in less-elevated circles. The problem remaining, then, is to understand the between-group result in terms of IQ, which might well be regarded simply as a race effect but for the fact that the prevalence rates turn out to be commensurate with IQ.

The answer probably lies in the fact that although education levels may order persons relatively with respect to IQ within each race, education does not control for IQ absolutely, and so it typically leaves uncontrolled a sizable Black–White IQ difference within levels of education, where a higher IQ could render persons less susceptible to unfounded beliefs in conspiracy. In contrast to performance outcomes, which always have practical consequences, the often theoretical and entirely verbal nature of opinions may serve as fertile ground for such a puzzling pattern of correlations. In contrast to Hochschild's results, suspiciousness (now unrelated to racial themes) has normally been found to correlate negatively, not positively, with education level, namely,  $-.33$ , controlling for both age and social desirability (Ranchor, Bouma, & Sanderman, 1996, Table 1). What ought to be surprising, perhaps, is that the model works at all for opinions.