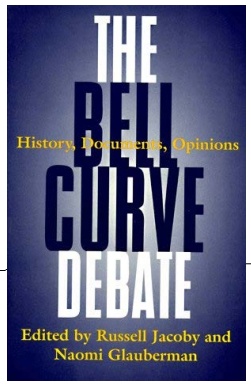
LIES, DAMNED LIES, AND STATISTICS

Leon J. Kamin



in Russell Jacoby and Naomi Glauberman, eds (1995)
The Bell Curve Debate. History, Documents, Opinions

WITHIN TWO MONTHS of its publication, 400,000 copies of *The Bell Curve* were in print, and Rep. Newt Gingrich of Georgia was elected Speaker of the House of Representatives. Those two events probably represent a correlation, rather than cause and effect, but the book and the congressman have a good deal in common. They let us know, up front, where they are coming from and where they are headed—which turn out to be the same place. We are going back, if they have their way, to a country familiar to Ebenezer Scrooge and Oliver Twist, and to a landscape dotted with orphanages and almshouses.

The publicity barrage with which the book was launched might suggest that *The Bell Curve* has something new to say; it doesn't. The authors, in this most recent eruption of the crude biological determinism that permeates the history of IQ testing, assert that scientific evidence demonstrates the existence of genetically determined differences in intelligence among social classes and races. They cite some 1,000 references from the social and biological sciences, and make a

Leon J. Kamin is professor of psychology at Northeastern University; he is author of *The Science and Politics of IQ*, and with R. C. Lewontin and Steven Rose of *Not in Our Genes*. This is an expanded version of a review that appeared in *Scientific American*, February 1995.

number of suggestions for changing social policies. The pretense is made that there is some logical, "scientific" connection between evidence culled from those cited sources and the authors' policy recommendations. Those policies would not be necessary or humane even if the cited evidence were valid. But I want to concentrate on what I regard as two disastrous failings of the book. First, the caliber of the data cited by Herrnstein and Murray is, at many critical points, pathetic—and their citations of those weak data are often inaccurate. Second, their failure to distinguish between correlation and causation repeatedly leads Herrnstein and Murray to draw invalid conclusions.

I'LL DEAL FIRST, at some length, with an especially troubling example of the quality of the data on which the authors rely. They begin their discussion of racial differences in IQ by assuring us that they "will undertake to confront all the tough questions squarely," and they caution us to "read carefully" as they "probe deeply into the evidence and its meaning." That tough, deep probing leads them to ask, "How Do African-Americans Compare with Blacks in Africa on Cognitive Tests?" Their reasoning is that low African-American IQ scores might be due either to a past history of slavery and discrimination or to genetic factors. Herrnstein and Murray evidently assume that blacks reared in colonial Africa have not been subjected to discrimination. Thus, if low IQ scores of African-Americans are a product of discrimination rather than genes, black Africans should have higher IQs than African-Americans; or so Herrnstein and Murray reason.

To answer the question they have posed, Herrnstein and Murray rely on the authority of Richard Lynn, described as "a leading scholar of racial and ethnic differences," from whose advice they have "benefited especially." They state that Lynn, who in 1991 reviewed eleven African IQ studies, "estimated the median black African IQ to be 75 . . . about ten points *lower* [emphasis added] than the current figure for American blacks." This means, they conclude, that the "special circumstances" of African-Americans cannot explain their low average IQ relative to whites. That leaves genetics free to explain the black-white difference.

But why do black Americans have higher scores than black Africans? Herrnstein and Murray, citing "Owen 1992" in support, write that "the IQ of 'coloured' students in South Africa—of mixed

racial background—has been found to be similar to that of American blacks." The implication is clear: the admixture of Caucasian and African genes, taking place in America as well as in South Africa, boosts "coloured" IQ some ten points above that of native Africans. But the claims made about African and coloured IQ levels cannot withstand critical scrutiny.

Lynn's 1991 paper describes a 1989 publication by Ken Owen as "the best single study of Negroid intelligence." That 1989 Owen study compared white, Indian, and black pupils on the "Junior Aptitude Tests"; no coloured pupils were tested. The mean "Negroid" IQ in this "best" study was, according to Lynn, 69. That was also, Lynn wrote, "around the median" IQ found in the eleven studies of "Negroid populations." He therefore suggested 70 as "the approximate mean for pure Negroids." I forbear to comment on Lynn's conclusion that half of all Africans are mentally retarded. (Herrnstein and Murray calculated the median of the eleven studies as 75, and took that value to represent average African IQ. I would like to believe that they added five IQ points to Lynn's estimate because they found 70 to be a ludicrously implausible figure, but I have no supporting evidence.)

But Owen did not in fact assign "IQs" to any of the groups he tested. He merely reported test score differences between groups in terms of standard deviation units. The IQ figure of 69 was concocted by Lynn out of those data. There is, as Owen made clear, no reason to suppose that the low test scores of blacks had much to do with genetics: "language played such an important role and the knowledge of English of the majority of black testees was so poor" that some of the tests proved to be "virtually unusable." The tests assumed that the Zulu pupils were familiar with such things as electrical appliances, microscopes, and "Western type of ladies' accessories." The original plan of research had been to draw the black sample from the same metropolitan areas as the whites and Indians. That was not possible, "owing to the unrest situation," so a black sample was obtained in KwaZulu.

In 1992 Owen reported on a sample of coloured students that had been added to the groups he had tested earlier. A footnote in *The Bell Curve* credits "Owen 1992" (the reference does not appear in the book's bibliography) as showing that South African coloured students

have an IQ "similar to that of American blacks"—i.e., about 85. That statement does not accurately characterize Owen's findings.

The test used by Owen in 1992 was the "nonverbal" Raven's Progressive Matrices, thought to be less culturally biased than most other IQ tests. He was now able to compare the performance of coloured students with that of the whites, blacks, and Indians in his 1989 study, since the earlier set of pupils had taken the Matrices as well as the Junior Aptitude Tests. The black pupils, recall, had poor knowledge of English, but Owen felt that instructions for the Matrices "are so easy that they can be explained with gestures."

In any event, Owen's 1992 paper again does not assign "IQs" to the pupils. The mean number of correct responses on the Matrices (out of a possible 60) is given for each group: 45 for whites, 42 for Indians, 37 for coloureds, and 28 for blacks. The test's developer, John Raven, always insisted that Progressive Matrices scores cannot be converted into IQs. The several standardizations of his test indicate only what raw score corresponds to what percentile score. The Matrices scores, unlike IQs, are not symmetrically distributed around their mean (no "bell curve" here). There is thus no meaningful way to convert an average of raw Matrices scores into an IQ, and no comparison with American black IQ is possible.

The percentile score to which the average raw score of a sample corresponds is not the same quantity as the average percentile score of the tested individuals. The skewed distribution of Matrices scores virtually guarantees that, in any sample with a reasonable spread of scores, those two quantities will differ considerably. Further, in Europe and America the average Matrices score has been increasing by about one standard deviation per generation; should one compare African scores to early (low) Western norms or to more recent (high) ones? These considerations did not prevent Lynn from converting average Matrices scores to percentile scores based on an unspecified Western standardization, and then, using the bell curve, transforming the percentile scores to "IQs."

To illustrate what Lynn has done, consider a small "thought experiment." We travel to Africa and give the Matrices test to a large number of children, all aged 13.5. Half of the children have raw scores of only thirteen correct answers, because they do not get the point and are merely guessing on the multiple choice test. The other half do get

the point, and all have raw scores of 56. The British standardization of 1979 indicates that those two raw scores fall at the 1st and 99th percentiles, respectively. Thus the average percentile score of the children is 50, corresponding to the exact center of the bell curve. The center of the bell curve, of course, implies an average IQ of 100. But Lynn would seize upon the fact that the average *raw* score was 34.5. That score corresponds to the 8th percentile in the standardization sample. Lynn, consulting the bell curve, would observe that the 8th percentile of a normal distribution corresponds to an IQ of 79, and would report that figure as the average Negroid IQ. Herrnstein and Murray would believe him; he is, after all, their expert.

The remaining studies cited by Lynn, and accepted as valid by Herrnstein and Murray, tell us little about African IQ, but do tell us something about Lynn's scholarship. Thus, one of the eleven entries in Lynn's table of the intelligence of "pure Negroids" indicates that 1,011 Zambians, tested with the Progressive Matrices, had a low average IQ of 75. The source for this quantitative claim is given as "Pons, 1974; Crawford Nutt, 1976." A. L. Pons did test 1,011 Zambian copper miners, whose average number of correct responses was 34. Pons reported on this work orally; his data were summarized in tabular form in a paper by D. H. Crawford-Nutt. Lynn took the Pons data from Crawford-Nutt's paper and converted the number of correct responses into a bogus average IQ of 75. But Lynn chose to ignore entirely the substance of Crawford-Nutt's paper, which reported that 228 black high school students in Soweto had an average of 45 correct responses on the Matrices—*higher* than the mean of 44 achieved by the same-aged white sample on whom the test's norms had been established, and well above the mean of Owen's coloured pupils. We should note that seven of the 11 studies which Lynn did choose to include in his "Negroid" table reported only average Matrices raw scores. The cited IQs are Lynn's inventions. The other studies used tests more clearly dependent on cultural content.

Lynn had earlier, in a 1978 paper, summarized six studies involving African pupils, most again based on the Matrices. The arbitrary "IQs" concocted by Lynn for those six studies ranged between 75 and 88, with a median of 84. There was almost no overlap between the studies selected for inclusion by Lynn in his 1978 and 1991 "summaries." Five of the studies cited in 1978 were omitted from Lynn's 1991 table, by which time African IQ had in his expert judgment plummeted to 69.

I will not mince words. Lynn's distortions and misrepresentations of the data constitute a truly venomous racism, combined with scandalous disregard for scientific objectivity. But to anybody familiar with Lynn's work and background, this comes as no surprise. Lynn is widely known to be an associate editor of the vulgarly racist journal *Mankind Quarterly;* his 1991 paper comparing the intelligence of "Negroids" and "Negroid-Caucasoid hybrids" appeared in its pages. He is a major recipient of financial support from the nativist and eugenically oriented Pioneer Fund. It is a matter of shame and disgrace that two eminent social scientists, fully aware of the sensitivity of the issues they address, take as their scientific tutor Richard Lynn, and accept uncritically his surveys of research. Murray, in a newspaper interview, asserted that he and Herrnstein had not inquired about the "antecedents" of the research they cite. "We used studies that exclusively, to my knowledge, meet the tests of scholarship." What tests of scholarship?

WHATEVER those tests might be, Herrnstein and Murray are not rigorous in applying them, even to the work of reputable scholars. To support their assertion that high IQ is a "preventative" against crime, they cite a Danish study based upon 1,400 boys. That study, they say, reported that sons whose fathers had a "prison record" were six times more likely to have a "prison record" themselves than were sons of fathers with "no police record of any sort." That fact is scarcely surprising, and is open to many different interpretations. But Herrnstein and Murray call attention to a further alleged fact. The sons of fathers with prison records can be regarded as being at "high risk" for imprisonment themselves. Among such high-risk sons, those who had "no police record at all" had IQs 13 points higher than those who "had a police record." Thus, according to Herrnstein and Murray, it is only the less bright among the sons of jailed criminals who themselves acquire police records.

That is not, however, what the Danish study reported. For a *father* to be classified as "severely criminal" he had to have received "at least one prison sentence." That one sentence placed his son into the high-risk category. For a *son* to be classified as "seriously criminal," two quite different definitions were employed by the researchers. To calculate the rate of "serious criminal behavior" among sons, the son—

like the father—need only have received one prison sentence. It was by use of that definition that high-risk sons were six times more likely to be seriously criminal (jailed) than were sons of fathers with no police record. But *to be included among the "seriously criminal" sons whose IQs were studied*, the son had to have received "at least one jail sentence plus an additional offense." With that new definition, the noncriminals among the high-risk sons had a higher IQ than the criminals; no such difference existed among low-risk sons. The 13-point IQ difference cited by Herrnstein and Murray is thus not simply between high-risk sons with and without "a police record."

There is no explanation given by the researchers as to why the definition of sons' criminality was changed when making the IQ analyses. The consequence of the change is that in calculating IQ scores, a son who is merely sentenced to prison for one rape is not counted as a criminal. To earn that designation he will have to rack up a parking ticket as well. To one steeped in the research literature of social science, a possible explanation for this unusual definition of criminality suggests itself. Perhaps if the definition of criminal for the IQ analyses were the same as that used for determining high risk, the data would not support the hypothesis tested by the research. That may not have been the case in this instance; but arbitrary post facto categorizing of data is not unheard of in science.

We should note in any event that most of the "additional offenses" which, when added to a jail sentence, qualified an at-risk son to be IQ tested could not have been very serious. Fully 57 percent of the 1,400 sons had such minor offenses on their records, in the absence of any jail sentence. Parking tickets and littering seem like reasonable candidates. What does a high IQ protect a high-risk Danish son against—committing rape or parking illegally? I don't know, and neither did Herrnstein and Murray.

HERE IS ANOTHER example of mis-citation in *The Bell Curve*, this time part of the effort to convince readers that blacks are less intelligent than whites. Herrnstein and Murray maintain that "smarter people process [information] faster than less smart people," and that reaction time, requiring "no conscious thought," indexes an underlying "neurologic processing speed . . . akin to the speed of the microprocessor in a computer." "Reaction time" is the time elapsing between onset of a sig-

nal light and a subject's lifting a finger to initiate a required response; "movement time" is the additional time needed to execute the response. Herrnstein and Murray report, "In modern studies, reaction time is correlated with the *g* factor in IQ tests. . . . Movement time is much less correlated with IQ. . . ." The cognitive processing, they explain, is measured by reaction time, while movement time measures "small motor skills." The work of Arthur Jensen is cited as follows: "The consistent result of many studies is that white reaction time is faster than black reaction time, but black movement time is faster than white movement time." White men can't jump, but they have faster computer chips inside their heads.

The cited Jensen paper (1993) presents data for blacks and whites, for both reaction and movement time, for three different "elementary cognitive tasks." The results are not, despite Herrnstein and Murray's contention, "consistent." Blacks are reported to have faster movement times on only two of the three tasks; and they have faster *reaction times* than whites on one task, "choice reaction time." Simple reaction time merely requires the subject to respond as quickly as possible to a given stimulus each time it occurs. Choice reaction time requires him/her to react differently to various stimuli as they are presented in an unpredictable order. Thus it is said to be more cognitively complex, and to require more processing, than simple reaction time. When Jensen first used reaction time in 1975 as a measure of racial differences in intelligence, he claimed that blacks and whites did not differ in simple reaction time, but that whites, with their higher intelligence, were faster in choice reaction time. He repeated this ludicrous claim incessantly, while refusing to make the raw data of his study available for inspection. Then, in a subsequent 1984 paper, he was unable to repeat his earlier finding in a new study described as "inexplicably inconsistent" with his 1975 results. Now, in the still newer 1993 study cited by Herrnstein and Murray, Jensen reports as "an apparent anomaly" that (once again!) blacks are slightly faster in choice reaction time than whites. Those swift couriers, Herrnstein and Murray, are not stayed from their appointed rounds by anomalies and inconsistencies. Two out of three is not conclusive. Why not make the series three out of five?

To anybody who has ever watched a professional basketball game, the idea that blacks are incapable of making quick choices about how to respond to complex and changing visual displays will not be very

convincing. How can scientists talk themselves into believing such a thing? But then, how can they talk themselves into believing that half of all Africans are mentally retarded? The answer to such questions doesn't require much thought. Murray, complaining to *The Wall Street Journal* that his book had been "blatantly misrepresented," blamed "the American preoccupation with race." Indeed.

I TURN NOW to a revealing example of Herrnstein and Murray's tendency to ignore the difference between a mere statistical association (correlation) and a cause-and-effect relationship. They lament that "private complaints about the incompetent affirmative-action hiree are much more common than scholarly examination of the issue." They proceed to a scholarly and public discussion of "teacher competency examinations." They report that such exams have had "generally beneficial effects," presumably by weeding out incompetent affirmative-action hirees. That positive view of standardized tests for teachers is not shared by those who argue that, since blacks tend to get lower scores, the tests are a way of eliminating competent black teachers. But Herrnstein and Murray assure us that "teachers who score higher on the tests have greater success with their students."

To support that claim they cite a single study by a couple of economists who analyzed data from a large number of North Carolina school districts. The researchers obtained average teacher test scores ("teacher quality") and average pupil failure rates for each district. They reported that a "1% increase in teacher quality . . . is accompanied by a 5% decline in the rate of failure of students." That is, there were fewer student failures in districts where teachers had higher test scores. But it does not follow from such a correlation that hiring teachers with higher test scores will reduce the rate of student failure. The same researchers found that "larger class size tends to lead to improved average [pupil] performance." Does it follow that increasing the pupil-to-teacher ratio will improve student performance? That policy recommendation might please many taxpayers, just as firing teachers with lower test scores would please some. But neither policy follows logically from the observed correlations.

To understand why, consider the following. The average proportion of black students across the school districts was 31 percent. Suppose—it does not stretch the limits of credibility—that there was a tendency

for black teachers (who have lower test scores) to work in districts with large proportions of black pupils (who have higher failure rates). That nonrandom assignment of teachers to classrooms would produce a correlation between teacher test scores and pupil failure rates—but one cannot then conclude that the teacher's test score has any causal relation to student failure. To argue that, we would have to show that for a group of black teachers (and for a separate group of white teachers) the teachers' test scores predicted the failure rates of their students. There was no such information available either to the original researchers or to Herrnstein and Murray.

What about the surprising finding that high pupil-teacher ratios are associated with good pupil performance? There's no way to be certain, but suppose deprived black children tended to be in small, de facto segregated rural schools, whereas more privileged whites were in larger classrooms. Would cramming more pupils into the rural schools promote academic excellence? There is a general and important lesson buried in this example: the arithmetical complexity of the multitude of correlations and logistic regressions stuffed into the Herrnstein-Murray volume does not elevate their status from mere associations to causes and effects.

THE CONFUSION between correlation and causation permeates the largest section of *The Bell Curve*, an interminable series of analyses of data gathered from the National Longitudinal Survey of Labor Market Experience of Youth (NLSY). Those data, not surprisingly, indicate that there is an association within each race between IQ and socioeconomic status (SES). Herrnstein and Murray labor mightily in an effort to show that low IQ is the cause of low SES, and not vice versa. Their argument is decked out in all the trappings of science—a veritable barrage of charts, graphs, tables, appendices, and appeals to statistical techniques that are unknown to many readers. But on close examination, this scientific emperor is wearing no clothes.

The NLSY survey included more than 12,000 youngsters who were aged fourteen to twenty-two when the continuing study began in 1979. The respondents and/or their parents at that time provided information about their educations, occupations, and income, and answered other questions about themselves. Those reports are the basis for classifying the childhood SES of the respondents. The teenagers also took the

Armed Forces Qualification Test, regarded by psychometricians as essentially an IQ test. As they have grown older, the respondents have provided more information about their own schooling, unemployment, poverty, marital status, childbearing, welfare dependency, criminality, parenting behavior, etc.

Herrnstein and Murray pick over these data, trying to show that it is overwhelmingly IQ—not childhood or adult SES—that determines worldly success and the moral praiseworthiness of one's social behaviors. But their dismissal of SES as a major factor rests ultimately on the self-reports of youngsters. That is not an entirely firm basis. I do not want to suggest that such self-reports are entirely unrelated to reality. We know, after all, that children from differing social class backgrounds do indeed differ in IQ; and in the NLSY study the young peoples' self-reports are correlated with the objective facts of their IQ scores. But comparing the predictive value of those self-reports to that of quantitative test scores is playing with loaded dice.

Further, the fact that self-reports are correlated with IQ scores is, like all correlations, ambiguous. For Herrnstein and Murray, the relation of their index of parental SES to the child's IQ means that high-SES parents—the "cream floating on the surface of American society"—have transmitted high quality genes to their offspring. But other interpretations are possible. Perhaps, for example, the kinds of people who get high test scores are precisely those who are vain enough to claim exaggerated social status for themselves. That tendency could artificially inflate correlations of IQ both with parental SES and with self-reports of success, distorting all tests of the relative predictive power of SES and IQ. That may seem far-fetched to some readers, but it is clearly a logical possibility. The choice between alternative interpretations of statistical associations cannot be based upon logic alone. There is thus plenty of elbow room for ideological bias in social science.

THE CORE of the Herrnstein-Murray message is phrased with a beguiling simplicity: "Putting it all together, success and failure in the American economy, and all that goes with it, are increasingly a matter of the genes that people inherit." The "increasing value of intelligence in the marketplace" brings "prosperity for those lucky enough to be intelligent." Income is a "family trait" because IQ, "a major predictor of income, passes on sufficiently from one generation to the

next to constrain economic mobility." Those at the bottom of the economic heap were unlucky when the IQ genes were passed out, and will remain there.

The correlations with which Herrnstein and Murray are obsessed are of course real: the children of day laborers are less likely than the children of stockbrokers to acquire fortunes or to go to college. They are more likely to be delinquent, to receive welfare, to have children outside of marriage, to be unemployed, and to have low-birth-weight babies. The children of laborers have lower average IQs than children of brokers, and so IQ is also related to all these phenomena. Herrnstein and Murray's intent is to convince us that low IQ causes poverty and its attendant evils—and not, as others might hold, vice versa.

For eight dense chapters they wrestle with data derived from the white respondents in the NLSY survey, attempting to disentangle the roles of IQ and of SES. They employ a number of quantitative tools, most prominently logistic regression—a technique that purports to specify what would happen if one variable is "held constant" while another variable is left free to vary. When SES is statistically "held constant" by Herrnstein and Murray, IQ remains related to all the phenomena described, in the obviously predictable direction. When IQ is held constant, the effect of SES is invariably reduced, usually very substantially, and sometimes eliminated.

There are a number of criticisms to be made of the ways in which Herrnstein and Murray analyze the data, and especially so when they later extend their analyses to include black and Hispanic youth. But for argument's sake, let us now suppose that their analyses are appropriate and accurate. We can also grant that, rightly or wrongly, disproportionate salaries and wealth accrue to those with high IQ scores. What then do the Herrnstein-Murray analyses tell us?

The SES of one's parents cannot in any direct sense "cause" one's IQ to be high or low. Family income, even if accurately reported, obviously cannot directly determine a child's performance on an IQ test. But income and the other components of an SES index can serve as rough indicators of the rearing environment to which a child has been exposed. With exceptions, a child of a well-to-do broker is likely to be exposed to book-learning earlier and more intensively than a child of a laborer. And extensive practice at reading and calculating does affect, very directly, one's IQ score. That is one plausible way of interpreting the statistical link between parental SES and a child's IQ.

The significant question is not whether the Herrnstein-Murray index of SES is more or less statistically associated with success than is their measure of IQ. Different SES measures, or different IQ tests, might substantially affect the results they obtained; other scholars, using other indices and tests, have gotten quite different results. The significant question is, why don't the children of laborers acquire the skills that are tapped by IQ tests?

Herrnstein and Murray answer that the children of the poor, like their laborer parents before them, have been born with poor genes. Armed with that conviction, they hail as "a great American success story" that after "controlling for IQ," ethnic and racial discrepancies in education and wages are "strikingly diminished." They reach this happy conclusion on the questionable basis of their regression analyses. But the data, even if true, would allow another reading. We can view it as a tragic failure of American society that so few black and low-SES children are lucky enough to be reared in environments that nurture development of the skills needed to obtain high IQ scores. For Herrnstein and Murray it is only fair that the race should go to the swift, and the swift are those blessed with good genes and high IQs. The conception that we live in a society that hobbles most of the racers at the starting line does not occur to them.

THE CONFIDENCE that Herrnstein and Murray appear to place in the ability of logistic regressions to interpret the social world seems excessive. To many readers that statistical procedure will be unknown, and thus beyond the reach of critical evaluation. That in turn will lead many to misunderstand the apparently simple charts scattered through the volume. The problem can be illustrated by a chart on page 322, captioned: "After controlling for IQ, blacks and Latinos have substantially higher probabilities than whites of being in a high-IQ occupation." The top panel of the chart indicates that "For a person of average age (29) before controlling for IQ," the probability of being in such an occupation is 5 percent for whites, 3 percent for blacks, and 3 percent for Latinos. The surface appearance, that blacks and Latinos are discriminated against, is misleading; logistic regression will demonstrate that.

The bottom panel of the chart shows that "for a person of average age and average IQ for people in high-IQ occupations (117)," the probability of being in such an occupation is 10 percent for whites, 26 percent for blacks, and 16 percent for Latinos. These adjusted proba-

bilities arise from using regression to "hold IQ constant," statistically, at the average value of NLSY respondents in high-IQ occupations (lawyers, doctors, et cetera). The insight afforded by the regression analysis is powerful. Those relatively rare blacks and Latinos who have IQs of 117, far from being discriminated against, are more likely than whites with the same high IQ to be in the high-income professions. Maybe affirmative action has degenerated into reverse racism.

The chart does not tell us the actual number, or actual proportions, of NLSY whites, blacks, and Latinos in the professions. The regression analysis has fitted a smooth curve through a cloud of actual data points. The probabilities in the chart have been read off from that idealized ("best-fitting") curve. We do not know how closely the curve fits the real data. We do know that since IQs as high as 117 are relatively rare, the curve at that point is based largely on extrapolating from the much more numerous data points at lower IQ levels. That extrapolation is pretty much an act of faith. How much so can be illustrated by a few simple and rough calculations.

There were 3,022 blacks in the total NLSY sample. The respondents were about equally distributed across eight different ages, with the same racial mix at all age levels. We can thus calculate that the sample of 29-year-olds (the top panel of the chart) contained about 378 blacks. The regression analysis predicts that 3 percent of them (about 11 people) should be in the professions. But it also tells us (the bottom panel) that among 29-year-old blacks with the necessary IQ (117 or higher), the probability of being in a profession skyrockets to 26 percent. We know that the average IQ of blacks in the NLSY sample was 86.7, with a standard deviation of 12.4. That enables us to calculate (the bell curve again) that 2.78 of the black 29-year-olds in the sample should have IQs of 117 or higher. The regression analysis informs us that fully 26 percent of those 2.78 blacks (0.72 of a black) are predicted to be in the professions. Murray is right; we are losing ground. Before the days of affirmative action, an entire token black was par for the course.

THE BELL CURVE's basic thesis is that "intelligence and its correlates—maturity, farsightedness, and personal competence—are important in keeping a person employed and in the labor force." That kind of theory is not new, and psychometricians are especially prone

to it. Raymond Cattell, described as "one of most [sic] illustrious psychometricians of his age," wrote during the Great Depression that "Unemployment—persistent unemployment—has unfortunately been regarded as a purely economic problem when in fact it is fundamentally a psychological one." The stress on psychological factors encourages Herrnstein and Murray to speculate on why, even if matched for IQ, blacks are more likely than whites to be unemployed. They raise "the possibility of ethnic differences in whatever other personal attributes besides IQ determine a person's ability to do well in the job market. We do not know whether ethnic groups differ on the average in these other ways. . . . We will not speculate further along these lines here." This tease encourages the reader to follow the authors into the locker room, where such speculations are routinely entertained. Professor Cattell was less shy about speculating in public. He wrote that the Negro race "has contributed practically nothing to social progress and culture (except in rhythm, sensitiveness to which is revealed by tests to be constitutionally better in the negro than the European)." Too bad that rhythm doesn't count for much in the job market.

Tests of cognitive ability, unlike tests of rhythm, are claimed by Herrnstein and Murray to be excellent predictors of "job productivity." Thus an employer concerned with the bottom line would do well to hire, no matter what the job, those applicants with high IQ test scores: "the smart busboy will be more productive than the less-smart busboy. . . ." But how do we measure the "productivity" of an employee? The vast majority of studies "validate" the predictive power of IQ tests by demonstrating that supervisors assign higher ratings to workers with high test scores. That fact, of course, tells us that supervisors think highly of workers with high test scores—most of whom share various traits (whiteness is one of them) with most supervisors. It does not necessarily tell us that high-IQ workers are more productive.

There is also an extensive research literature which demonstrates that workers with high IQs possess more "job knowledge," as assessed by written multiple-choice tests. High-IQ workers are also more likely to pass written qualifying examinations given at the end of training courses for particular jobs. But again, these facts do not demonstrate that—once on the job—high-IQ workers are really more productive.

There have been some studies, many conducted by the military, in which the criterion for job productivity has involved actual work samples, or "hands-on" tests. Maier and Hiatt, in a technical report cited by Herrnstein and Murray, explain that "hands-on job performance tests have intrinsic validity because of their high fidelity to the skills required to perform job tasks. . . . [they] are the benchmark measure for evaluating the job relatedness of surrogate measures of job performance, such as written tests, ratings, and grades."

With an understanding of how psychologists measure job productivity, we can now follow Herrnstein and Murray as they grapple with the problem of whether experience on the job can "make up for less intelligence." They conclude that "the difference in productivity associated with differences in intelligence diminishes only slowly and partially. Often it does not diminish at all. The cost of hiring less intelligent workers may last as long as they stay on the job." To arrive at this bleak conclusion, they cite only two studies (both in the military) which used work samples or hands-on tests. Their description of one study is false; their description of the other study is accurate, but incomplete.

Herrnstein and Murray assert that Schmidt et al. studied armor repairmen, armor crewmen, supply specialists, and cooks "extending out to five years of experience and using three different measures of job performance." They indicate that the researchers found high-IQ workers to begin at higher levels, and to continue to outstrip low-IQ workers by the same amount, in all jobs, for all measures, for five years. That much is basically true, but it obscures an important fact. In all measures—work samples, job knowledge tests, and supervisory ratings—both high- and low-IQ workers improved steadily with experience. Thus, in work sample scores, a low-IQ worker after two years was about as productive as a high-IQ worker after one year of experience. Facts of that sort are not irrelevant to the productive utilization of "human capital."

But more; despite Herrnstein and Murray's claim that the study extended out to five years, 194 of the 1,457 workers had had more than five years of experience. The work sample scores of such highly experienced low-IQ workers had completely caught up to those of equally experienced high-IQ workers! The supervisory ratings of the experienced low-IQ workers were actually higher than those of high-IQ work-

ers, although a substantial gap remained in "job knowledge" tests. These embarrassments were explained away by the study's authors with an appeal to "a fluke of sampling error," and an assertion that "findings in the highest experience group are suspect."

The second military study cited by Herrnstein and Murray is that of Maier and Hiatt. That study was described, accurately enough, as finding that a difference favoring high-IQ workers persisted over time when "job knowledge" was the criterion, but disappeared when a work sample was the measure. The data in fact indicated that, for both ground radio repairers and automotive mechanics, high-IQ workers initially outscored low-IQ workers on both hands-on and written tests. But after four or five years of experience, the low-IQ workers actually did better on the hands-on test than those with high IQs! On the written test of "job knowledge," low-IQ workers showed no sign whatever of catching up to the superior multiple-choice testing skills of their high-IQ betters. Maier and Hiatt concluded that the military's IQ test was "a valid predictor of job performance as measured by hands-on tests," but that the content validity of hands-on tests "is sensitive to job experience." That is a psychometrician's way of saying that after a few years on the job the correlation between IQ and worker productivity was actually slightly negative.

This military research, I think, has a genuine and deep meaning. The kinds of people who don't do well on standardized tests have some trouble catching on to job requirements in the early going; but with experience their actual work performance catches up to that of their more academically talented peers. Their problem appears to be that even when they are doing the job excellently, they have no "job knowledge." They don't *know* how to do the job, they just *do* it; or at least they can't write down what they do know. That, in the view of Herrnstein and Murray, is sufficient reason to consign them to unemployment.

In the world of *The Bell Curve*, the importance and the explanatory power of IQ are ubiquitous. Before the advent of IQ tests, "gossip about who in the tribe was cleverer" was "a topic of conversation around the fire since fires, and conversation, were invented." Among Bushmen of the Kalahari, "the best hunters score above their tribal average on IQ tests." Faced with the choice, it is "better to be born smart [than] rich."

Herrnstein and Murray note that among blue-collar workers who tell researchers that they have dropped out of the labor force because of physical disability or injury, low IQ is common. Why? "An answer leaps to mind: The smarter you are, the less likely that you will have accidents." That answer leapt to mind before the thought that low-IQ workers, in minimum wage jobs, have little incentive to remain in the labor force. Dull young women lack the "foresight and intelligence" to understand that the welfare system offers them a bad deal. Welfare might be a bad deal for Herrnstein and Murray, but I am not so sure that single mothers on welfare haven't figured out *their* odds pretty accurately.

A low-IQ woman is likely to have a low-birth-weight baby because she "never registers the simple and ubiquitous lessons about taking care of herself" when pregnant. Her problem is not that she has no prenatal care; it is that she has "difficulty in connecting cause and effect." People who have low IQs, according to *The Bell Curve*, commit crimes because, lacking foresight, the threat of prison does not deter them; further, they cannot "understand why robbing someone is wrong." Then what is to be made of the fact that although "very dull" young males are stopped by the police, booked for an offense, and convicted of an offense less often than "normal" males, they are nevertheless jailed more than twice as often? "It may be . . . that they are less competent in getting favorable treatment from the criminal justice system. The data give us no way to tell." Perhaps not, but some hints are available. There is no doubt that O. J. Simpson is "competent"; but his ability to hire high-priced lawyers is not irrelevant to the treatment he will receive from the criminal justice system.

THE BELL CURVE, near its closing tail, contains two chapters concerned with affirmative action, in higher education and in the workplace. To read those chapters is to hear the second shoe drop. The rest of the book, I believe, was written merely as a prelude to its assault on affirmative action. The vigor of the attack is astonishing.

Affirmative action "cannot survive public scrutiny." It is based on "the explicit assumption that ethnic groups do not differ in . . . abilities." Hiring and promotion procedures "that are truly fair . . . will produce . . . racial disparities," and "employers are using double standards for black and white applicants . . . because someone or something . . . is making them do so. . . ." The "degradation of intellectual require-

ments" in recruiting police has affected "police performance on the street." We learn that a veteran of the Washington, D.C., police force has heard "about people in the academy who could not read or write." And a former instructor saw "people diagnosed as borderline retarded graduate from the police academy." These anecdotes take their place among the politically potent folk tales about welfare queens driving Cadillacs.

Herrnstein and Murray contribute to the genre by describing a black student who "it was reported, received a straight grant of $85,000, plus $10,000 in annual travel budgets, from one of Harvard's competitors in minority recruiting." Their cited source for this tale is the *Harvard University Gazette*. The account in that journal quotes a Harvard admissions officer as having learned, through "an informal poll," of an African-American student who was offered "a grant of $85,000 over four years, plus an additional $10,000 each summer for travel and *research*" [emphasis added]. When I asked that admissions officer for specific details, he replied that the principle of confidentiality prevented him from answering. He did, however, cite as a relevant "minority scholarship" the Angier B. Duke scholarships awarded by Duke University. Inquiry at Duke established that these are not "minority scholarships." They are full tuition, four-year scholarships awarded each year, without regard to need, to the sixteen most outstanding applicants to Duke. They include support to travel for summer study at Oxford University in England. This seems the likely source of *The Bell Curve*'s saga about the Willie Horton of the Ivy League.

Now, at long last, Herrnstein and Murray let it all hang out: "affirmative action, in education and the workplace alike, is leaking a poison into the American soul." Having examined the American condition at the close of the twentieth century, these two philosopher-kings conclude, "It is time for America once again to try living with inequality, as life is lived. . . ." This kind of sentiment, I imagine, lay behind the conclusion of *New York Times* columnist Bob Herbert that "the book is just a genteel way of calling somebody a nigger." Herbert is right. The book has nothing to do with science.

WITH *THE BELL CURVE* proper behind us, I want now to consider one of the more pernicious effects of its publication. The enormously successful marketing of the book by its publisher and by the

American Enterprise Institute has served to legitimize as "scholarship" overtly racist works which only a year or two ago were widely regarded as outside the mainstream of academic respectability. The *New York Times* science reporter, Malcolm Browne, appropriately chose to review the Herrnstein and Murray volume together with recent books by J. Philippe Rushton (*Race, Evolution, and Behavior*) and by Seymour Itzkoff (*The Decline of Intelligence in America*). Browne, in lumping the books together, assured readers of the *Times* that "the government or society that persists in sweeping their subject matter under the rug will do so at its peril." We can only hope, perhaps naively, that exposure to the light of critical scrutiny might have some antiseptic effect against the scholarship of writers like Rushton and Itzkoff.

Rushton has written that human evolution has produced three major races—Mongoloids, Caucasoids, and Negroids. These races are said to differ, in the same rank ordering, with respect to a large number of correlated physical and behavioral traits, all related to "reproductive strategies." Those traits—all of which Rushton believes to be encoded in the genes of the different races—include intelligence, brain size, penis size, nurturing one's young, frequency of sexual intercourse, number of offspring, law-abidingness, sexual hormone levels, the tendency to have low-birth-weight babies, and altruism. For these and other traits Negroids are said to be at one end of a continuum, far removed from Caucasoids. Mongoloids are at the other end of the continuum, but close to Caucasoids. The Rushton portrait of Negroids—stupid, small brains, big penises, sexually licentious, criminal, spawning lots of low-birth-weight babies for whom they will not care—strikes a responsive chord in America; David Duke was almost elected governor of Louisiana.

Herrnstein and Murray grant that "Rushton paints with a broad brush," but write of his "detailed and convincing empirical reports of the race differences," and declare that his "work is not that of a crackpot or a bigot. . . . As science, there is nothing wrong with Rushton's work in principle. . . ." I'll mention just a couple of the empirical details that Herrnstein and Murray found convincing. Rushton asserts that blacks have larger penises than whites. Presumably this scholar's understanding of human sexuality includes the belief that big penises are more likely to engage in intercourse and to produce babies than

are small penises. To demonstrate that blacks have big penises, Rushton cited just two sources—some casual observations by an anonymous French army surgeon in Africa writing in 1898, and some unpublished data from Kinsey's study of American sexual behavior. The volunteer male subjects in the Kinsey study were asked to measure their own penises. The proportion of black subjects complying with that request was significantly smaller than the proportion of whites. The few blacks who did comply—scarcely a random sample of blacks—claimed slightly larger penis sizes than the many whites who responded.

To demonstrate that black genes produce unbridled sexual behavior, as well as big genitals, Rushton reported that a significantly higher proportion of black than of white interviewees had told Kinsey that the female partner tended to have more than one orgasm per act of intercourse. To assert this as a fact Rushton—unknown to his readers—had to lump together the responses of male and female interviewees, which had been tabled separately by Kinsey. The actual data were that 18 percent of black males, but only 8 percent of black females (!), claimed that the female had multiple orgasms; among whites, the proportion making that claim was 9 percent of both males and females. The data as published by Rushton indicated simply that 13 percent of blacks and 9 percent of whites reported multiple female orgasms. That evidently qualifies in the eyes of Herrnstein and Murray as a "detailed and convincing report of the race differences." What would Rushton have to write before *The Bell Curve*'s authors would conclude that he is a crackpot or a bigot?

Predictably, Rushton's theorizing has excited the prurient interest of Herrnstein and Murray's psychometric expert, Richard Lynn. "The high rate of sexual activity in Negroids," Lynn has suggested, may be caused by a high level of the male sex hormone, testosterone. The "crucial supporting evidence" for the notion that blacks have an oversupply of testosterone is the fact that "Negroids have higher rates of cancer of the prostate than Caucasoids . . . an important determinant of cancer of the prostate is the level of testosterone." The chain of reasoned evidence is: prostate cancer is caused by testosterone; blacks tend to have prostate cancer; therefore blacks must have lots of testosterone; the abundance of testosterone makes blacks sexually active; that causes them to produce lots of babies, for whom they will

not provide, and who will become criminals and/or welfare cases. It's all in the genes.

This train of reasoning can be headed off at the pass. To show that testosterone causes prostate cancer (a view not widely shared in medical circles), Lynn cites a paper by Ahluwalia et al. That paper, Lynn writes, reported "higher levels of testosterone in patients with prostatic cancer than in healthy controls." That claim, like Rushton's claim about multiple female orgasms, does not quite tell the whole truth. Ahluwalia et al. reported that black prostate patients in the United States had higher testosterone levels than did control subjects. But among blacks in Nigeria, control subjects had higher testosterone levels than did prostate patients! Testosterone appears to cause prostate cancer in America, while protecting Nigerians from the same affliction.

What about the next claim, that blacks are more prone than whites to develop prostate cancer? That again is partially true—but not in the sweeping racial sense that Lynn intends. Lynn reprints some age-standardized incidence rates of prostate cancer for "Negroids" and "Caucasoids" in seven American cities. Those statistics and others had been gathered by the International Union Against Cancer. There was variation from city to city, but in each case African-Americans had about twice the incidence of whites. The highest white rate was 59.7 per 100,000 population, in Hawaii (Lynn erroneously attributes that rate to Hawaiian "Negroids"); the lowest black rate was 72.1, in New Orleans.

The paper from which Lynn copied (or tried to copy) those figures contains other relevant statistics. The rate in Senegal was 4.3—the lowest rate, except for Japan and Shanghai, among the thirty-odd countries for which data were given. The rates in Jamaica and (then) Rhodesia were 28.6 and 32.3—still far below the rates of both black and white Americans. Follow-up studies by the International Union reported a rate of 9.7 in Nigeria. In the Cape Province of South Africa, the rate for whites was a low 23.2; for Bantus it was 19.2, and for Africans in Natal 23.2. The facts are well known to every serious scholar concerned with prostate cancer: American blacks have an alarmingly higher rate of prostate cancer than American whites, but black Africans have a much lower rate than either American blacks or whites. These facts do not lend themselves to the racist interpretations advanced by Herrnstein and Murray's psychometric expert, Richard Lynn. To admit Lynn and Rushton into the scientific mainstream—I'll say it bluntly—is a

betrayal of science. To say this out loud is not to advocate what Malcolm Browne describes as a "shroud of censorship imposed upon scientists and scholars by pressure groups." It is a simple defense of truth and integrity in science. Herrnstein and Murray's defense of Rushton's racist claptrap—"we expect that time will tell whether it is right or wrong in fact"—is couched in the tones of moderation and reason. In my view both the work and its defense are contemptible.

THE WORK by Itzkoff, and its echoes in *The Bell Curve,* could (and should) have been written seventy years ago; in fact, it was. Browne summarizes Itzkoff's views with entire accuracy: "the least intelligent, least educable, poorest, most politically apathetic and abusive contingent of the population is reproducing faster than the smart, rich, politically active and nurturing contingent . . . this has fueled a dysgenic trend: America's collective smartness is being diluted, gravely endangering the nation's ability to compete economically." Herrnstein and Murray similarly bemoan the alleged propensity for the cognitively least able to reproduce excessively; and worse yet, once more guided by "Richard Lynn's computations," they conclude that America's "immigrants in the 1980s came from ethnic groups that have [IQ] scores significantly below the white average. . . ."

The same phenomena had seized the attention of Carl C. Brigham in 1923. Brigham, convinced that excessive breeding by the lower classes must produce a decline in "American intelligence," analyzed the mental test scores of foreign-born draftees into the American army during World War I. Those data indicated that immigrants from southern and eastern Europe, and Russia ("our army sample of immigrants from Russia is at least one half Jewish"), had appallingly low IQs. Brigham advocated, and Congress enacted, laws to minimize the proportion of immigrants admitted from southern and eastern Europe. He warned that "racial admixture" in America "is infinitely worse than that faced by any European country today, for we are incorporating the negro into our racial stock, while all of Europe is comparatively free from this taint. . . . The decline of American intelligence will be more rapid than the decline of the intelligence of European national groups, owing to the presence here of the negro." Brigham looked forward to "the prevention of the continued propagation of defective strains in the present population." He, and these views, were not outside the

mainstream of psychological science in 1923; Brigham went on to become secretary both of the American Psychological Association and of the College Entrance Examination Board, where he developed the Scholastic Aptitude Test.

Malcolm Browne, commenting on the books by Herrnstein and Murray, by Rushton, and by Itzkoff, suggests that "the authors . . . may have softened their agendas somewhat to parry the expected fury of liberal critics, fellow academics and hostile mobs. . . . it is hard to believe that these writers would oppose a eugenically motivated program designed to influence patterns of reproduction." The notion that these writers labor under a "shroud of censorship" imposed by "pressure groups," or that the lavishly endowed American Enterprise Institute trembles before the expected fury of liberal critics, academics, and hostile mobs, seems out of touch with what is really happening in America. What, other than "a eugenically motivated program," is the Herrnstein-Murray recommendation to end welfare aid to unmarried mothers with dependent children?

The specter of dysgenesis has haunted psychometrics since its inception; no material facts are capable of dislodging that specter. Raymond Cattell, then in England, wrote an entire book in 1937 on *The Fight for Our National Intelligence.* The fact that intelligence, measured by IQ, was inherited was self-evident; the only opposition to that view came from "enemies of democracy" and "people primarily political in outlook." While Hitler swept Europe before him, Cattell—that "most illustrious" psychometrician—explained that since "intelligence tests point to significant differences between races," it was "people racially in a temporarily awkward tactical position" who opposed the findings of the IQ testers.

The tendency of the lower classes to breed excessively, and of the upper classes to restrict their fertility, must surely—unless counteracted—lead to a decline in "national intelligence." Cattell, joined by virtually all the leading psychometricians of the time, confidently predicted that national surveys would show a decline in average IQ of some 1.5 points per decade. When national surveys showed instead that there had been an *increase* in average IQ over time, psychometricians concluded that the test used in the surveys (the Stanford-Binet) was an imperfect measure of "innate" intelligence. Whatever imperfect tests might indicate, actual intelligence *had* to have declined.

That psychometric tradition of heads-I-win-tails-you-lose has been carried forward intact by Herrnstein and Murray. They acknowledge that James Flynn has demonstrated that across the world intelligence as measured by IQ tests has been increasing dramatically over time. Thus an average contemporary youngster, taking an IQ test that had been standardized twenty years ago, would have a considerably higher than average IQ score. Perhaps, Herrnstein and Murray suggest, "Improved health, education, and childhood interventions may hide the demographic effects. . . . Whatever good things we can accomplish with changes in the environment would be that much more effective if they did not have to fight a demographic head wind." Their conviction that "something worth worrying about is happening to the cognitive capital of the country" is unshakable. Imagine the heights that America could scale if a Ph.D. in social science were a prerequisite for the production of offspring! With environmental advantages working exclusively upon such splendid raw material, no head winds would delay our arrival at Utopia. And we would sell more autos to the Japanese.

That is the kind of brave new world toward which *The Bell Curve* points. Whether or not our country moves in that direction depends upon our politics, not upon science. To pretend, as Herrnstein and Murray do, that the 1,000-odd items in their bibliography provide a "scientific" basis for their reactionary politics may be a clever political tactic, but it is a disservice to and abuse of science. That should be clear even to those scientists (I am not one of them) who are comfortable with Herrnstein and Murray's politics. We owe it to our fellow citizens to explain that the reception of their book had nothing to do either with its scientific merit or the novelty of its message.