Note: What follows is a fully annotated version of the article that appears in the print edition of the September 2005 issue of COMMENTARY.

The Inequality Taboo

Charles Murray

When the late Richard Herrnstein and I published *The Bell Curve* eleven years ago, the furor over its discussion of ethnic differences in IQ was so intense that most people who have not read the book still think it was about race. Since then, I have deliberately not published anything about group differences in IQ, mostly to give the real topic of *The Bell Curve*—the role of intelligence in reshaping America's class structure—a chance to surface.

The Lawrence Summers affair last January made me rethink my silence. The president of Harvard University offered a few mild, speculative, off-the-record remarks about innate differences between men and women in their aptitude for high-level science and mathematics, and was treated by Harvard's faculty as if he were a crank. The typical news story portrayed the idea of innate sex differences as a renegade position that reputable scholars rejected.

It was depressingly familiar. In the autumn of 1994, I had watched with dismay as *The Bell Curve*'s scientifically unremarkable statements about black IQ were successfully labeled as racist pseudoscience. At the opening of 2005, I watched as some scientifically unremarkable statements about male-female differences were successfully labeled as sexist pseudoscience.

The Orwellian disinformation about innate group differences is not wholly the media's fault. Many academics who are familiar with the state of knowledge are afraid to go on the record. Talking publicly can dry up research funding for senior professors and can cost assistant professors their jobs. But while the public's misconception is understandable, it is also getting in the way of clear thinking about American social policy.

Good social policy can be based on premises that have nothing to do with scientific truth. The premise that is supposed to undergird all of our social policy, the founders' assertion of an unalienable right to liberty, is not a falsifiable hypothesis. But specific policies based on premises that conflict with scientific truths about human beings tend not to work. Often they do harm.

One such premise is that the distribution of innate abilities and propensities is the same across different groups. The statistical tests for

uncovering job discrimination assume that men are not innately different from women, blacks from whites, older people from younger people, homosexuals from heterosexuals, Latinos from Anglos, in ways that can legitimately affect employment decisions. Title IX of the Educational Amendments of 1972 assumes that women are no different from men in their attraction to sports. Affirmative action in all its forms assumes there are no innate differences between any of the groups it seeks to help and everyone else. The assumption of no innate differences among groups suffuses American social policy. That assumption is wrong.

When the outcomes that these policies are supposed to produce fail to occur, with one group falling short, the fault for the discrepancy has been assigned to society. It continues to be assumed that better programs, better regulations, or the right court decisions can make the differences go away. That assumption is also wrong.

Hence this essay. Most of the following discussion describes reasons for believing that some group differences are intractable. I shift from "innate" to "intractable" to acknowledge how complex is the interaction of genes, their expression in behavior, and the environment. "Intractable" means that, whatever the precise partitioning of causation may be (we seldom know), policy interventions can only tweak the difference at the margins.

I will focus on two sorts of differences: between men and women and between blacks and whites. Here are three crucial points to keep in mind as we go along:

1. The differences I discuss involve means and distributions. In all cases, the variation within groups is greater than the variation between groups. On psychological and cognitive dimensions, some members of both sexes and all races fall everywhere along the range. One implication of this is that genius does not come in one color or sex, and neither does any other human ability. Another is that a few minutes of conversation with individuals you meet will tell you much more about them than their group membership does.

2. Covering both sex differences and race differences in a single, non-technical article, I had to leave out much in the print edition of this article. This online version is fully annotated and includes extensive supplementary material.

3. The concepts of "inferiority" and "superiority" are inappropriate to group comparisons. On most specific human attributes, it is possible to specify a continuum running from "low" to "high," but the results cannot be combined into a score running from "bad" to "good." What is the best score on a continuum measuring aggressiveness? What is the relative importance of verbal skills versus, say, compassion? Of spatial skills

versus industriousness? The aggregate excellences and shortcomings of human groups do not lend themselves to simple comparisons. That is why the members of just about every group can so easily conclude that they are God's chosen people. All of us use the weighting system that favors our group's strengths.1

II

The technical literature documenting sex differences and their biological basis grew surreptitiously during feminism's heyday in the 1970's and 1980's. By the 1990's, it had become so extensive that the bibliography in David Geary's pioneering *Male, Female* (1998) ran to 53 pages.2 Currently, the best short account of the state of knowledge is Steven Pinker's chapter on gender in *The Blank Slate* (2002).3

Rather than present a telegraphic list of all the differences that I think have been established, I will focus on the narrower question at the heart of the Summers controversy: as groups, do men and women differ innately in characteristics that produce achievement at the highest levels of accomplishment? I will limit my comments to the arts and sciences.

Since we live in an age when students are likely to hear more about Marie Curie than about Albert Einstein, it is worth beginning with a statement of historical fact: women have played a proportionally tiny part in the history of the arts and sciences.4 Even in the 20th century, women got only 2 percent of the Nobel Prizes in the sciences—a proportion constant for both halves of the century—and 10 percent of the prizes in literature. The Fields Medal, the most prestigious award in mathematics, has been given to 44 people since it originated in 1936. All have been men.

The historical reality of male dominance of the greatest achievements in science and the arts is not open to argument. The question is whether the social and legal exclusion of women is a sufficient explanation for this situation, or whether sex-specific characteristics are also at work.

Mathematics offers an entry point for thinking about the answer. Through high school, girls earn better grades in math than boys, but the boys usually do better on standardized tests.5 The difference in means is modest, but the male advantage increases as the focus shifts from means to extremes. In a large sample of mathematically gifted youths, for example, seven times as many males as females scored in the top percentile of the SAT mathematics test.6 We do not have good test data on the male-female ratio at the top one-hundredth or top one-thousandth of a percentile, where first-rate mathematicians are most likely to be found, but collateral evidence suggests that the male advantage there continues to increase, perhaps exponentially.7

Evolutionary biologists have some theories that feed into an explanation for the disparity. In primitive societies, men did the hunting, which often took them far from home. Males with the ability to recognize landscapes from different orientations and thereby find their way back had a survival advantage. Men who could process trajectories in three dimensions—the trajectory, say, of a spear thrown at an edible mammal—also had a survival advantage.8 Women did the gathering. Those who could distinguish among complex arrays of vegetation, remembering which were the poisonous plants and which the nourishing ones, also had a survival advantage. Thus the logic for explaining why men should have developed elevated three-dimensional visuospatial skills and women an elevated ability to remember objects and their relative locations—differences that show up in specialized tests today.9

Perhaps this is a just-so story.10 Why not instead attribute the results of these tests to socialization? Enter the neuroscientists. It has been known for years that, even after adjusting for body size, men have larger brains than women. Yet most psychometricians conclude that men and women have the same mean IQ (although debate on this issue is growing).11 One hypothesis for explaining this paradox is that three-dimensional processing absorbs the extra male capacity. In the last few years, magnetic-resonance imaging has refined the evidence for this hypothesis, revealing that parts of the brain's parietal cortex associated with space perception are proportionally bigger in men than in women.12

What does space perception have to do with scores on math tests?13 Enter the psychometricians, who demonstrate that when visuospatial ability is taken into account, the sex difference in SAT math scores shrinks substantially.14

Why should the difference be so much greater at the extremes than at the mean? Part of the answer is that men consistently exhibit higher variance than women on all sorts of characteristics, including visuospatial abilities, meaning that there are proportionally more men than women at both ends of the bell curve.15 Another part of the answer is that someone with a high verbal IQ can easily master the basic algebra, geometry, and calculus that make up most of the items in an ordinary math test. Elevated visuospatial skills are most useful for the most difficult items.16 If males have an advantage in answering those comparatively few really hard items, the increasing disparity at the extremes becomes explicable.

Seen from one perspective, this pattern demonstrates what should be obvious: there is nothing inherent in being a woman that precludes high math ability. But there remains a distributional difference in male and female characteristics that leads to a larger number of men with high visuospatial skills. The difference has an evolutionary rationale, a physiological basis, and a direct correlation with math scores.

Now put all this alongside the historical data on accomplishment in the arts and sciences. In test scores, the male advantage is most pronounced in the most abstract items. Historically, too, it is most pronounced in the most abstract domains of accomplishment.17

In the humanities, the most abstract field is philosophy—and no woman has been a significant original thinker in any of the world's great philosophical traditions. In the sciences, the most abstract field is mathematics, where the number of great women mathematicians is approximately two (Emmy Noether definitely, Sonya Kovalevskaya maybe). In the other hard sciences, the contributions of great women scientists have usually been empirical rather than theoretical, with leading cases in point being Henrietta Leavitt, Dorothy Hodgkin, Lise Meitner, Irène Joliot-Curie, and Marie Curie herself.

In the arts, literature is the least abstract and by far the most rooted in human interaction; visual art incorporates a greater admixture of the abstract; musical composition is the most abstract of all the arts, using neither words nor images. The role of women has varied accordingly. Women have been represented among great writers virtually from the beginning of literature, in East Asia and South Asia as well as in the West. Women have produced a smaller number of important visual artists, and none that is clearly in the first rank. No female composer is even close to the first rank. Social restrictions undoubtedly damped down women's contributions in all of the arts, but the pattern of accomplishment that did break through is strikingly consistent with what we know about the respective strengths of male and female cognitive repertoires.

Women have their own cognitive advantages over men, many of them involving verbal fluency and interpersonal skills. If this were a comprehensive survey, detailing those advantages would take up as much space as I have devoted to a particular male advantage. But, sticking with my restricted topic, I will move to another aspect of male-female differences that bears on accomplishment at the highest levels of the arts and sciences: motherhood.

Regarding women, men, and babies, the technical literature is as unambiguous as everyday experience would lead one to suppose. As a rule, the experience of parenthood is more profoundly life-altering for women than for men. Nor is there anything unique about humans in this regard. Mammalian reproduction generally involves much higher levels of maternal than paternal investment in the raising of children.18 Among humans, extensive empirical study has demonstrated that women are more attracted to children than are men, respond to them more intensely

on an emotional level, and get more and different kinds of satisfactions from nurturing them. Many of these behavioral differences have been linked with biochemical differences between men and women.19

Thus, for reasons embedded in the biochemistry and neurophysiology of being female, many women with the cognitive skills for achievement at the highest level also have something else they want to do in life: have a baby. In the arts and sciences, forty is the mean age at which peak accomplishment occurs, preceded by years of intense effort mastering the discipline in question.20 These are precisely the years during which most women must bear children if they are to bear them at all.

Among women who have become mothers, the possibilities for high-level accomplishment in the arts and sciences shrink because, for innate reasons, the distractions of parenthood are greater. To put it in a way that most readers with children will recognize, a father can go to work and forget about his children for the whole day. Hardly any mother can do this, no matter how good her day-care arrangement or full-time nanny may be. My point is not that women must choose between a career and children, but that accomplishment at the extremes commonly comes from a single-minded focus that leaves no room for anything but the task at hand.21 We should not be surprised or dismayed to find that motherhood reduces the proportion of highly talented young women who are willing to make that tradeoff.

Some numbers can be put to this observation through a study of nearly 2,000 men and women who were identified as extraordinarily talented in math at age thirteen and were followed up 20 years later.22 The women in the sample came of age in the 1970's and early 1980's, when women were actively socialized to resist gender stereotypes. In many ways, these talented women did resist. By their early thirties, both the men and women had become exceptional achievers, receiving advanced degrees in roughly equal proportions. Only about 15 percent of the women were full-time housewives. Among the women, those who did and those who did not have children were equally satisfied with their careers.

And yet. The women with careers were four-and-a-half times more likely than men to say they preferred to work fewer than 40 hours per week. The men placed greater importance on "being successful in my line of work" and "inventing or creating something that will have an impact," while the women found greater value in "having strong friendships," "living close to parents and relatives," and "having a meaningful spiritual life." As the authors concluded, "these men and women appear to have constructed satisfying and meaningful lives that took somewhat different forms."23 The different forms, which directly influence the likelihood that men will dominate at the extreme levels of achievement, are consistent with a constellation of differences between men and women that have biological roots.

I have omitted perhaps the most obvious reason why men and women differ at the highest levels of accomplishment: men take more risks, are more competitive, and are more aggressive than women.[24] The word "testosterone" may come to mind, and appropriately. Much technical literature documents the hormonal basis of personality differences that bear on sex differences in extreme and venturesome effort, and hence in extremes of accomplishment—and that bear as well on the male propensity to produce an overwhelming proportion of the world's crime and approximately 100 percent of its wars. But this is just one more of the ways in which science is demonstrating that men and women are really and truly different, a fact so obvious that only intellectuals could ever have thought otherwise.

III

Turning to race, we must begin with the fraught question of whether it even exists, or whether it is instead a social construct. The Harvard geneticist Richard Lewontin originated the idea of race as a social construct in 1972, arguing that the genetic differences across races were so trivial that no scientist working exclusively with genetic data would sort people into blacks, whites, or Asians. In his words, "racial classification is now seen to be of virtually no genetic or taxonomic significance."[25]

Lewontin's position, which quickly became a tenet of political correctness, carried with it a potential means of being falsified. If he was correct, then a statistical analysis of genetic markers would not produce clusters corresponding to common racial labels.

In the last few years, that test has become feasible, and now we know that Lewontin was wrong.[26] Several analyses have confirmed the genetic reality of group identities going under the label of race or ethnicity.[27] In the most recent, published this year, all but five of the 3,636 subjects fell into the cluster of genetic markers corresponding to their self-identified ethnic group.[28] When a statistical procedure, blind to physical characteristics and working exclusively with genetic information, classifies 99.9 percent of the individuals in a large sample in the same way they classify themselves, it is hard to argue that race is imaginary.

Homo sapiens actually falls into many more interesting groups than the bulky ones known as "races."[29] As new findings appear almost weekly, it seems increasingly likely that we are just at the beginning of a process that will identify all sorts of genetic differences among groups, whether the groups being compared are Nigerian blacks and Kenyan blacks, lawyers and engineers, or Episcopalians and Baptists. At the moment, the differences that are obviously genetic involve diseases (Ashkenazi Jews and Tay-Sachs disease, black Africans and sickle-cell anemia, Swedes

and hemochromatosis). As time goes on, we may yet come to understand better why, say, Italians are more vivacious than Scots.

Out of all the interesting and intractable differences that may eventually be identified, one in particular remains a hot button like no other: the IQ difference between blacks and whites. What is the present state of our knowledge about it?

There is no technical dispute on some of the core issues. In the aftermath of *The Bell Curve*, the American Psychological Association established a task force on intelligence whose report was published in early 1996.30 The task force reached the same conclusions as *The Bell Curve* on the size and meaningfulness of the black-white difference. Historically, it has been about one standard deviation31 in magnitude among subjects who have reached adolescence;32 cultural bias in IQ tests does not explain the difference; and the tests are about equally predictive of educational, social, and economic outcomes for blacks and whites. However controversial such assertions may still be in the eyes of the mainstream media, they are not controversial within the scientific community.

The most important change in the state of knowledge since the mid-1990's lies in our increased understanding of what has happened to the size of the black-white difference over time. Both the task force and *The Bell Curve* concluded that some narrowing had occurred since the early 1970's. With the advantage of an additional decade of data, we are now able to be more precise: (1) The black-white difference in scores on educational achievement tests has narrowed significantly. (2) The black-white convergence in scores on the most highly "*g*-loaded" tests—the tests that are the best measures of cognitive ability—has been smaller, and may be unchanged, since the first tests were administered 90 years ago.

With regard to the difference in educational achievement, the narrowing of scores on major tests occurred in the 1970's and 80's. In the case of the SAT, the gaps in the verbal and math tests as of 1972 were 1.24 and 1.26 standard deviations respectively.33 By 1991, when the gaps were smallest (they have risen slightly since then), those numbers had dropped by .37 and .35 standard deviations.

The National Assessment of Educational Progress (NAEP), which is not limited to college-bound students, is preferable to the SAT for estimating nationally representative trends, but the story it tells is similar.34 Among students ages nine, thirteen, and seventeen, the black-white differences in math as of the first NAEP test in 1973 were 1.03, 1.29, and 1.24 standard deviations respectively. For nine-year-olds, the difference hit its all-time low of .73 standard deviations in 2004, a drop of .30 standard deviations.

But almost all of that convergence had been reached by 1986, when the gap was .78 standard deviations. For thirteen-year-olds, the gap dropped by .45 standard deviations, reaching its low in 1986. For seventeen-year-olds, the gap dropped by .52 standard deviations, reaching its low in 1990.

In the reading test, the comparable gaps for ages nine, thirteen, and seventeen as of the first NAEP test in 1971 were 1.12, 1.17, and 1.25 standard deviations. Those gaps had shrunk by .38, .62, and .68 standard deviations respectively at their lowest points in 1988.[35] They have since remained effectively unchanged.

An analysis by Larry Hedges and Amy Nowell uses a third set of data, examining the trends for high-school seniors by comparing six large data bases from different time periods from 1965 to 1992. The black-white difference on a combined measure of math, vocabulary, and reading fell from 1.18 to .82 standard deviations in that time, a reduction of .36 standard deviations.[36]

So black and white academic achievement converged significantly in the 1970's and 1980's, typically by more than a third of a standard deviation, and since then has stayed about the same.[37] What about convergence in tests explicitly designed to measure IQ rather than academic achievement?[38] The ambiguities in the data leave two defensible positions. The first is that the IQ difference is about one standard deviation, effectively unchanged since the first black-white comparisons 90 years ago. The second is that harbingers of a narrowing difference are starting to emerge. I cannot settle the argument here, but I can convey some sense of the uncertainty.

The case for an unchanged black-white IQ difference is straightforward. If you take all the black-white differences on IQ tests from the first ones in World War I up to the present, there is no statistically significant downward trend. Of course the results vary, because tests vary in the precision with which they measure the general mental factor (g) and samples vary in their size and representativeness. But results continue to center around a black-white difference of about 1.0 to 1.1 standard deviations through the most recent data.[39]

The case for a reduction has two important recent results to work with. The first is from the 1997 re-norming of the Armed Forces Qualification Test (AFQT), which showed a black-white difference of .97 standard deviations.[40] Since the typical difference on paper-and-pencil IQ tests like the AFQT has been about 1.10 standard deviations, the 1997 results represent noticeable improvement.[41] The second positive result comes from the 2003 standardization sample for the Wechsler Intelligence Scale

for Children (WISC-IV), which showed a difference of .78 standard deviations, as against the 1.0 difference that has been typical for individually administered IQ tests.[42]

One cannot draw strong conclusions from two data points. Those who interpret them as part of an unchanging overall pattern can cite another recent result, from the 2001 standardization of the Woodcock-Johnson intelligence test. In line with the conventional gap, it showed an overall black-white difference of 1.05 standard deviations and, for youths aged six to eighteen, a difference of .99 standard deviations.[43]

There is more to be said on both sides of this issue, but nothing conclusive.[44] Until new data become available, you may take your choice. If you are a pessimist, the gap has been unchanged at about one standard deviation. If you are an optimist, the IQ gap has decreased by a few points, but it is still close to one standard deviation. The clear and substantial convergence that occurred in academic tests has at best been but dimly reflected in IQ scores, and at worst not reflected at all.

Whether we are talking about academic achievement or about IQ, are the causes of the black-white difference environmental or genetic? Everyone agrees that environment plays a part. The controversy is about whether biology is also involved.

It has been known for many years that the obvious environmental factors such as income, parental occupation, and schools explain only part of the absolute black-white difference and none of the relative difference. Black and white students from affluent neighborhoods are separated by as large a proportional gap as are blacks and whites from poor neighborhoods.[45] Thus the most interesting recent studies of environmental causes have worked with cultural explanations instead of socioeconomic status.[46]

One example is *Black American Students in an Affluent Suburb: A Study of Academic Disengagement* (2003) by the Berkeley anthropologist John Ogbu, who went to Shaker Heights, Ohio, to explore why black students in an affluent suburb should lag behind their white peers.[47] Another is *Black Rednecks and White Liberals* (2005) by Thomas Sowell, who makes the case that what we think of as the dysfunctional aspects of urban black culture are a legacy not of slavery but of Southern and rural white "cracker" culture.[48] Both Ogbu and Sowell describe ingrained parental behaviors and student attitudes that must impede black academic performance. These cultural influences often cut across social classes.

From a theoretical standpoint, the cultural explanations offer fresh ways of looking at the black-white difference at a time when the standard socioeconomic explanations have reached a dead end. From a practical

standpoint, however, the cultural explanations point to a cause of the black-white difference that is as impervious to manipulation by social policy as causes rooted in biology. If there is to be a rapid improvement, some form of mass movement with powerful behavioral consequences would have to occur within the black community. Absent that, the best we can hope for is gradual cultural change that is likely to be measured in decades.

This brings us to the state of knowledge about genetic explanations. "There is not much direct evidence on this point," said the American Psychological Association's task force dismissively, "but what little there is fails to support the genetic hypothesis."[49] Actually, there is no direct evidence at all, just a wide variety of indirect evidence, almost all of which the task force chose to ignore.[50]

As it happens, a comprehensive survey of that evidence, and of the objections to it, appeared this past June in the journal *Psychology, Public Policy, and Law*. There, J. Philippe Rushton and Arthur Jensen co-authored a 60-page article entitled "Thirty Years of Research on Race Differences in Cognitive Ability."[51] It incorporates studies of East Asians as well as blacks and whites and concludes that the source of the black-white-Asian difference is 50- to 80-percent genetic. The same issue of the journal includes four commentaries, three of them written by prominent scholars who oppose the idea that any part of the black-white difference is genetic.[52] Thus, in one place, you can examine the strongest arguments that each side in the debate can bring to bear.

Rushton and Jensen base their conclusion on ten categories of evidence that are consistent with a model in which both environment and genes cause the black-white difference and inconsistent with a model that requires no genetic contribution.[53] I will not try to review their argument here, or the critiques of it. All of the contributions can be found on the Internet, and can be understood by readers with a grasp of basic statistical concepts.[54]

For those who consider it important to know what percentage of the IQ difference is genetic, a methodology that would do the job is now available. In the United States, few people classified as black are actually of 100-percent African descent (the average American black is thought to be about 20-percent white).[55] To the extent that genes play a role, IQ will vary by racial admixture. In the past, studies that have attempted to test this hypothesis have had no accurate way to measure the degree of admixture, and the results have been accordingly muddy.[56] The recent advances in using genetic markers solves that problem. Take a large sample of racially diverse people, give them a good IQ test, and then use genetic markers to create a variable that no longer classifies people as "white" or "black," but along a continuum. Analyze the variation in IQ scores according to that continuum. The results would be close to

dispositive.[57]

None of this is important for social policy, however, where the issue is not the source of the difference but its intractability. Much of the evidence reviewed by Rushton and Jensen bears on what we can expect about future changes in the black-white IQ difference. My own thinking on this issue is shaped by the relationship of the difference to a factor I have already mentioned—"*g*"—and to the developing evidence for *g*'s biological basis.

When you compare black and white mean scores on a battery of subtests, you do not find a uniform set of differences; nor do you find a random assortment. The size of the difference varies systematically by type of subtest. Asked to predict which subtests show the largest difference, most people will think first of ones that have the most cultural content and are the most sensitive to good schooling. But this natural expectation is wrong. Some of the largest differences are found on subtests that have little or no cultural content, such as ones based on abstract designs.

As long ago as 1927, Charles Spearman, the pioneer psychometrician who discovered *g*, proposed a hypothesis to explain the pattern: the size of the black-white difference would be "most marked in just those [subtests] which are known to be saturated with *g*."[58] In other words, Spearman conjectured that the black-white difference would be greatest on tests that were the purest measures of intelligence, as opposed to tests of knowledge or memory.

A concrete example illustrates how Spearman's hypothesis works. Two items in the Wechsler and Stanford-Binet IQ tests are known as "forward digit span" and "backward digit span." In the forward version, the subject repeats a random sequence of one-digit numbers given by the examiner, starting with two digits and adding another with each iteration. The subject's score is the number of digits that he can repeat without error on two consecutive trials. Digits-backward works exactly the same way except that the digits must be repeated in the opposite order.

Digits-backward is much more *g*-loaded than digits-forward. Try it yourself and you will see why. Digits-forward is a straightforward matter of short-term memory. Digits-backward makes your brain work much harder.[59]

The black-white difference in digits-backward is about twice as large as the difference in digits-forward.[60] It is a clean example of an effect that resists cultural explanation. It cannot be explained by differential educational attainment, income, or any other socioeconomic factor. Parenting style is irrelevant. Reluctance to "act white" is irrelevant.

Motivation is irrelevant. There is no way that any of these variables could systematically encourage black performance in digits-forward while depressing it in digits-backward in the same test at the same time with the same examiner in the same setting.[61]

In 1980, Arthur Jensen began a research program for testing Spearman's hypothesis. In his book *The* g *Factor* (1998), he summarized the results from seventeen independent sets of data, derived from 149 psychometric tests. They consistently supported Spearman's hypothesis.[62] Subsequent work has added still more evidence.[63] Debate continues about what the correlation between *g*-loadings and the size of the black-white difference means, but the core of Spearman's original conjecture, that a sizable correlation would be found to exist, has been confirmed.[64]

During the same years that Jensen was investigating Spearman's hypothesis, progress was also being made in understanding *g*. For decades, psychometricians had tried to make *g* go away. Confident that intelligence must be more complicated than a single factor, they strove to replace *g* with measures of uncorrelated mental skills. They thereby made valuable contributions to our understanding of intelligence, which really does manifest itself in different ways and with different profiles, but getting rid of *g* proved impossible. No matter how the data were analyzed, a single factor kept dominating the results.[65]

By the 1980's, the robustness and value of *g* as an explanatory construct were broadly accepted among pyschometricians, but little was known about its physiological basis.[66] As of 2005, we know much more. It is now established that *g* is by far the most heritable component of IQ.[67] A variety of studies have found correlations between *g* and physiological phenomena such as brain-evoked potentials, brain pH levels, brain glucose metabolism, nerve-conduction velocity, and reaction time.[68] Most recently, it has been determined that a highly significant relationship exists between *g* and the volume of gray matter in specific areas of the frontal cortex, and that the magnitude of the volume is under tight genetic control.[69] In short, we now know that *g* captures something in the biology of the brain.

So Spearman's basic conjecture was correct—the size of the black-white difference and *g*-loadings are correlated—and *g* represents a biologically grounded and highly heritable cognitive resource. When those two observations are put together, a number of characteristics of the black-white difference become predictable, correspond with phenomena we have observed in data, and give us reason to think that not much will change in the years to come.[70]

One implication is that black-white convergence on test scores will be

greatest on tests that are least g-loaded. Literacy is the obvious example: people with a wide range of IQ's can be taught to read competently, and it is the reading test of the NAEP in which convergence has reached its closest point (.55 standard deviations in the 1988 test). More broadly, the confirmation of Spearman's hypothesis explains why the convergence that has occurred on academic achievement tests has not been matched on IQ tests.

A related implication is that the source of the black-white difference lies in skills that are hardest to change. Being able to repeat many digits backward has no value in itself. It points to a valuable underlying mental ability, in the same way that percentage of fast-twitch muscle fibers points to an underlying athletic ability. If you were to practice reciting digits backward for a few days, you could increase your score somewhat, just as training can improve your running speed somewhat. But in neither case will you have improved the underlying ability.[71] As far as anyone knows, *g* itself cannot be coached.

The third implication is that the "Flynn effect" will not close the black-white difference. I am referring here to the secular increase in IQ scores over time, brought to public attention by James Flynn.[72] The Flynn effect has been taken as a reason for thinking that the black-white difference is temporary: if IQ scores are so malleable that they can rise steadily for several decades, why should not the black-white difference be malleable as well?[73]

But as the Flynn effect has been studied over the last decade, the evidence has grown, and now seems persuasive, that the increases in IQ scores do not represent significant increases in *g*.[74] What the increases do represent—whether increases in specific mental skills or merely increased test sophistication—is still being debated. But if the black-white difference is concentrated in *g* and if the Flynn effect does not consist of increases in *g*, the Flynn effect will not do much to close the gap. A 2004 study by Dutch scholars tested this question directly. Examining five large databases, the authors concluded that "the nature of the Flynn effect is qualitatively different from the nature of black-white differences in the United States," and that "the implications of the Flynn effect for black-white differences appear small."[75]

These observations represent my reading of a body of evidence that is incomplete, and they will surely have to be modified as we learn more. But taking the story of the black-white IQ difference as a whole, I submit that we know two facts beyond much doubt. First, the conventional environmental explanation of the black-white difference is inadequate. Poverty, bad schools, and racism, which seem such obvious culprits, do not explain it. Insofar as the environment *is* the cause, it is not the sort of environment we know how to change, and we have tried every practical remedy that anyone has been able to think of. Second, regardless of one's

reading of the competing arguments, we are left with an IQ difference that has, at best, narrowed by only a few points over the last century. I can find nothing in the history of this difference, or in what we have learned about its causes over the last ten years, to suggest that any faster change is in our future.

IV

Elites throughout the West are living a lie, basing the futures of their societies on the assumption that all groups of people are equal in all respects. Lie is a strong word, but justified. It is a lie because so many elite politicians who profess to believe it in public do not believe it in private. It is a lie because so many elite scholars choose to ignore what is already known and choose not to inquire into what they suspect. We enable ourselves to continue to live the lie by establishing a taboo against discussion of group differences.

The taboo is not perfect—otherwise, I would not have been able to document this essay—but it is powerful. Witness how few of Harvard's faculty who understood the state of knowledge about sex differences were willing to speak out during the Summers affair. In the public-policy debate, witness the contorted ways in which even the opponents of policies like affirmative action frame their arguments so that no one can accuse them of saying that women are different from men or blacks from whites. Witness the unwillingness of the mainstream media to discuss group differences without assuring readers that the differences will disappear when the world becomes a better place.

The taboo arises from an admirable idealism about human equality. If it did no harm, or if the harm it did were minor, there would be no need to write about it. But taboos have consequences.

The nature of many of the consequences must be a matter of conjecture because people are so fearful of exploring them.[76] Consider an observation furtively voiced by many who interact with civil servants: that government is riddled with people who have been promoted to their level of incompetence because of pressure to have a staff with the correct sex and ethnicity in the correct proportions and positions. Are these just anecdotes? Or should we be worrying about the effects of affirmative action on the quality of government services?[77] It would be helpful to know the answers, but we will not so long as the taboo against talking about group difference prevails.

How much damage has the taboo done to the education of children? Christina Hoff Sommers has argued that willed blindness to the different developmental patterns of boys and girls has led many educators to see boys as aberrational and girls as the norm, with pervasive damage to the

way our elementary and secondary schools are run.78 Is she right? Few have been willing to pursue the issue lest they be required to talk about innate group differences. Similar questions can be asked about the damage done to medical care, whose practitioners have only recently begun to acknowledge the ways in which ethnic groups respond differently to certain drugs.79

How much damage has the taboo done to our understanding of America's social problems? The part played by sexism in creating the ratio of males to females on mathematics faculties is not the ratio we observe but what remains after adjustment for male-female differences in high-end mathematical ability. The part played by racism in creating different outcomes in black and white poverty, crime, and illegitimacy is not the raw disparity we observe but what remains after controlling for group characteristics. For some outcomes, sex or race differences nearly disappear after a proper analysis is done. For others, a large residual difference remains.80 In either case, open discussion of group differences would give us a better grasp on where to look for causes and solutions.

What good can come of raising this divisive topic? The honest answer is that no one knows for sure. What we do know is that the taboo has crippled our ability to explore almost any topic that involves the different ways in which groups of people respond to the world around them—which means almost every political, social, or economic topic of any complexity.

Thus my modest recommendation, requiring no change in laws or regulations, just a little more gumption. Let us start talking about group differences openly—all sorts of group differences, from the visuospatial skills of men and women to the vivaciousness of Italians and Scots. Let us talk about the nature of the manly versus the womanly virtues. About differences between Russians and Chinese that might affect their adoption of capitalism. About differences between Arabs and Europeans that might affect the assimilation of Arab immigrants into European democracies. About differences between the poor and non-poor that could inform policy for reducing poverty.

Even to begin listing the topics that could be enriched by an inquiry into the nature of group differences is to reveal how stifled today's conversation is. Besides liberating that conversation, an open and undefensive discussion would puncture the irrational fear of the male-female and black-white differences I have surveyed here. We would be free to talk about other sexual and racial differences as well, many of which favor women and blacks, and none of which is large enough to frighten anyone who looks at them dispassionately.

Talking about group differences does not require any of us to change our politics. For every implication that the Right might seize upon (affirmative-action quotas are ill-conceived), another gives fodder to the Left (innate group differences help rationalize compensatory redistribution by the state).[81] But if we do not need to change our politics, talking about group differences obligates all of us to renew our commitment to the ideal of equality that Thomas Jefferson had in mind when he wrote as a self-evident truth that all men are created equal. Steven Pinker put that ideal in today's language in *The Blank Slate*, writing that "Equality is not the empirical claim that all groups of humans are interchangeable; it is the moral principle that individuals should not be judged or constrained by the average properties of their group."[82]

Nothing in this essay implies that this moral principle has already been realized or that we are powerless to make progress. In elementary and secondary education, many outcomes are tractable even if group differences in ability remain unchanged. Dropout rates, literacy, and numeracy are all tractable. School discipline, teacher performance, and the quality of the curriculum are tractable. Academic performance within a given IQ range is tractable. The existence of group differences need not and should not discourage attempts to improve schooling for millions of American children who are now getting bad educations.

In university education and in the world of work, overall openness of opportunity has been transformed for the better over the last half-century. But the policies we now have in place are impeding, not facilitating, further progress. Creating double standards for physically demanding jobs so that women can qualify ensures that men in those jobs will never see women as their equals. In universities, affirmative action ensures that the black-white difference in IQ in the population at large is brought onto the campus and made visible to every student. The intentions of their designers notwithstanding, today's policies are perfectly fashioned to create separation, condescension, and resentment—and so they have done.

The world need not be that way. Any university or employer that genuinely applied a single set of standards for hiring, firing, admitting, and promoting would find that performance across different groups really is distributed indistinguishably. But getting to that point nationwide will require us to jettison an apparatus of laws, regulations, and bureaucracies that has been 40 years in the making. That will not happen until the conversation has opened up. So let us take one step at a time. Let us stop being afraid of data that tell us a story we do not want to hear, stop the name-calling, stop the denial, and start facing reality.

CHARLES MURRAY *is the W.H. Brady Scholar in Freedom and Culture at the American Enterprise Institute. His previous contributions to* COMMENTARY*, available online, include* "The Bell Curve *and Its*

Critics" (May 1995, with a subsequent exchange in the August 1995 issue).

Notes

My thanks go to Michael Ashton, Thomas Bouchard, Gregory Carey, Christopher DeMuth, David Geary, Linda Gottfredson, Arthur Jensen, John Loehlin, David Lubinski, Kevin McGrew, Richard McNally, Derek Neal, Steven Pinker, Philip Roth, Philippe Rushton, Sally Satel, Christina Hoff Sommers, Hua Tang, Marley Watkins, Lawrence Weiss, and James Q. Wilson for responding to questions or commenting on drafts. Their appearance on this list does not imply their endorsement of anything in the essay.

---

1 If you think this is mushy nonjudgmentalism, try a thought experiment: Suppose that a pill exists that, if all women took it, would give them exactly the same mean and variance on every dimension of human functioning as men—including all the ways in which women now surpass men. How many women would want all women to take it? Or suppose that the pill, taken by all blacks, would give them exactly the same mean and variance on every dimension of human functioning as whites—including all the ways in which blacks now surpass whites. How many blacks would want all blacks to take it? To ask such questions is to answer them: hardly anybody. Few want to trade off the unique virtues of their own group for the advantages that another group may enjoy.

Sometimes these preferences for one's own group are rational, sometimes not. I am proud of being Scots-Irish, for example, even though the Scots-Irish group means for violence, drunkenness, and general disagreeableness seem to have been far above those of other immigrant groups. But the Scots-Irish made great pioneers—that's the part of my heritage that I choose to value. A Thai friend gave me an insight into this human characteristic many years ago when I remarked that Thais were completely undefensive about Westerners despite the economic backwardness of Thailand in those days. My friend explained why. America has wealth and technology that Thailand does not have, he acknowledged, just as the elephant is stronger than a human. "But," he said with a shrug, "who wants to be an elephant?" None of us wants to be an elephant and, from the perspective of our own group, every other group has something of the elephant about it. All of us are right, too.

2 Geary (1998).

3 Pinker (2002). A non-technical book-length treatment is Rhoads (2004). Halpern (2000) and Kimura (1999) are good one-volume discussions of cognitive differences between the sexes. An up-to-date summary of neuro-physiological findings about sex differences in the brain appeared in last May's *Scientific American*, Cahill (2005). Baron-Cohen (2003) is an ambitious attempt to tie together known sex differences into an overall theory. Those who want to compare these accounts with defenses of the no-innate-differences position can look at Valian (1999) and a set of essays weighted toward social explanations of math differences in Gallagher and Kaufman (2005).

4 My discussion of women and accomplishment in the arts and sciences is in Murray (2003): 265–293. For a complementary discussion, see Simonton (1999): chapter 6.

5 For the story on grades, see Kimball (1989). For a review of the literature on male-female differences in means and methods of mathematical processing, see Geary, Saults, Liu *et al.* (2000). For discussions of sex stereotyping, see Brown and Josephs (1999), Stipek and Gralinski (1991), and several of the essays in Gallagher and Kaufman (2005).

6 This ratio is based on the percentages of boys and girls from Talent Search who later, as high-school students, got the top possible score in the SAT-Math (12.7 percent of males and 1.9 percent of females, given in Lubinski, Benbow, Shea *et al.*, 2001). Julian Stanley, who has been associated with Talent Search for many years, is said to have asserted in an interview that the male:female ratio among such students has dropped to about 3 to 1. I have not been able to locate the interview or any data substantiating that ratio. In any case, here is a reminder: currently, the 800 top score in the SAT-Math is only about 2.6 standard deviations above the mean—that is, it includes about one in 200 test-takers. This is nowhere close to the extreme right end of the bell curve from which top mathematicians are drawn.

7 Nyborg (in press) finds a sex difference in the general mental ability *g*, not just in spatial skills, and evidence that the male advantage increases exponentially as distance from the mean increases.

8 For a review of studies about sex differences in throwing ability, see Geary (1998): 213–14, 284–85. For a presentation of the evolutionary explanation, see Jones, Braithwaite, and Healy (2003) and Kimura (1999): 11–30. It has also been argued that spatial skills were an advantage in tool-making. See Wynn, Tierson, and Palmer (1996).

9 Geary (1998): 286–90; Kimura (1999): 43–66.

10 A continuing problem for evolutionary biology is the accusation that its scholars observe human characteristics today and work backward into a rationale that fits. But a sex difference in visuospatial abilities is found in many other animals besides humans, always favoring males—which gives good reason for thinking that in this case we are observing something more than a just-so story. See Jones, Braithwaite, and Healy (2003).

11 For a review of the evidence that male and female IQ is the same, see Jensen (1998): 536–42. The underlying problem is that the subtests in IQ tests have been developed and normed in ways that tend to push male and female IQs toward the same mean IQ (for example, items that show a large sex difference are usually discarded). For the evidence that men have a higher mean IQ than women, see Ankney (1992), Lynn (1999), Lynn and Irwing (2004), and Nyborg (in press).

12 See Goldstein, Seidman, Horton *et al.* (2001) and the interpretation of those findings in Cahill (2005). This is far from a settled issue. Research into the neurophysiology of sex differences is exploring a variety of trails. For example, Gron, Spitzer, Tomczak *et al.* (2000) discovered that men and women activate different parts of the brain when they are working out navigation tasks, and do so in patterns consistent with the proposition that navigation is cognitively more difficult for women. Consistent evidence also links the size of brain regions with level of capability (Cahill 2005). This relationship between specific parts of the brain and capability holds at an aggregate level as well: IQ is correlated with brain size (adjusted for body size). The relationship of brain size to IQ has often been derided (e.g., Gould 1981), and indeed brain size was a problematic measure when it had to be based on skull size or post-mortem data. But magnetic resonance imaging (MRI) studies of brain size have ended the uncertainty about the existence of its relationship with IQ. For meta-analyses of MRI and other *in vivo* studies, see Jensen (1998): 147, which puts the correlation between brain size and IQ at about .40, and McDaniel (2005), which puts it at about .33.

13 E.g., Johnson (1984), Casey, Nuttall, Pezaris *et al.* (1995), and Geary, Saults, Liu *et al.* (2000). There has been dispute on this point. Friedman (1995) argues that performance in math tests is more strongly related to verbal ability than to visuospatial abilities. Royer, Tronsky, Chan *et al.* (1999) present evidence that the real source of the male advantage is faster retrieval of arithmetic facts from long-term memory. A third line of argument has been that the apparent male advantage is actually mediated by IQ (e.g., Linn and Peterson 1985). Geary, Saults, Liu *et al.* (2000) controlled for IQ and found that both visuospatial abilities and the computational advantage found by Royer, Tronsky, Chan *et al.* (1999) were at work.

14 Casey, Nuttall, Pezaris *et al.* (1995).

15 Pinker (2002): 344–45.

16 Visuospatial skills are helpful across the entire range of items (see Geary, Saults, Liu *et al.* 2000), but good verbal skills can substitute in solving the less difficult items.

17 For a more detailed presentation of the evidence about the pattern of female accomplishment in the arts and sciences, see Murray (2003): 265–69.

18 Geary (1998): 20–28, 97–120.

19 For an analysis of sex differences in nurturing, written by a committed feminist who is also a scientist (an anthropologist), see Hrdy (1999). For a short review of studies on the importance of children and of the biological sources of nurturing differences, see Rhoads (2004): 190–222.

20 Simonton (1984): chapter 6.

21 Ochse (1990), Simonton (1994): chapter 5.

22 Benbow, Lubinski, Shea *et al.* (2000).

23 Ibid., 479. The figures in the text combine the data reported for two separate cohorts.

24 For a meta-analysis of sex differences in risk-taking, see Byrnes, Miller, and Schafer (1999). For a discussion of the role of testosterone, see J.M. Dabbs and M.G. Dabbs (2000).

25 Lewontin (1972).

26 For a technical description of what has been labeled "Lewontin's fallacy," see Edwards (2003). For a nontechnical statement of how the understanding of this issue has been changing, see Leroi (2005).

27 Studies incorporating some variant of this type of analysis include Bamshad, Wooding, Watkins *et al.* (2003), Bowcock, Ruiz-Linares, Romfohrde *et al.* (1994), Calafell, Shuster, Speed *et al.* (1998), Mountain and Cavalli-Sforza (1997), Rosenberg, Pritchard, Weber *et al.* (2002), and Stephens, Schneider, Tanguay *et al.* (2001).

28 Tang, Quertermous, Rodriguez *et al.* (2005). The self-identified ethnic groups consisted of non-Hispanic black, non-Hispanic white, East Asian, and Hispanic. The statistical procedure was cluster analysis. The algorithms in cluster analysis are not trying to find groupings that correspond to any pre-identified characteristic of the people in the

sample—that is, the researchers did not use any information about the physical characteristics that humans use to identify ethnicity. Cluster analysis simply looks for interrelationships among the genetic markers that identify statistically distinct entities.

29 In Tang, Quertermous, Rodriguez *et al.* (2005), "Hispanic" corresponded to a cluster, even though no one thinks of "Hispanic" as a race. People do not need to belong to different races, conventionally defined, to be genetically distinct.

30 Neisser, Boodoo, Bouchard *et al.* (1996).

31 The standard deviation is a statistic that (simplified) expresses the average difference of all the scores from the mean. More precisely, the standard deviation is calculated by squaring the deviation from the mean for each score, summing all those squared deviations, finding the mean of that sum, then taking the square root of the result. Given a normal distribution—a bell curve—someone who is one standard deviation above the mean is at the 84th percentile. Two standard deviations above the mean put that person at the 98th percentile. IQ tests are normed to have a mean of 100 and a standard deviation of 15.

32 The black-white difference emerges as early as IQ can be tested, but the gap is usually smaller in pre-adolescence. Among pre-schoolers, the gap can be just a few IQ points. Why does it increase with age? One obvious hypothesis is inferior schooling—e.g., Fryer and Levitt (2004). But black children attending excellent schools also fall behind their white counterparts, as discussed subsequently in the text and in note 14. The alternative explanation is that the heritability of IQ increases with age for people of all races, and this is reflected in black IQ scores in adolescence and adulthood. See Jensen (1998): 178.

33 My analysis of its annual College-Bound Seniors report, distributed as printed material prior to 1996 and available online from 1996 onward.

A word about the method of calculating the difference. When comparing scores from two groups, the preferred method is to divide the difference in the two scores by the pooled standard deviations of the two groups. The equation is

$$\frac{(\overline{X}_a - \overline{X}_b)}{\sqrt{(N_a\sigma_a^2 + N_b\sigma_b^2)/(N_a + N_b)}}$$

where N is the sample size, X is the sample mean, σ is the standard deviation, and the subscripts a and b denote each group. When the black-

white difference for a specific test is reported subsequently in the text, this equation has been used to compute it.

34 The Long Term Trend Study with consistent data for the NAEP from the early 1970's through 2004 is now available in mathematics and reading for students tested at ages nine, thirteen, and seventeen.

35 For nine-year-olds, the gap in reading scores expressed as points was smaller in 2004 (26 points) than in 1988 (29 points), but the difference in standard deviations was fractionally larger (.76 standard deviations in 2004 as compared with .74 in 1988).

36 Hedges and Nowell (1998): 154.

37 I will venture a prediction that a variety of academic achievement measures in elementary and secondary school will soon show renewed convergence because of the No Child Left Behind Act, which puts schools under intense pressure to teach to the test in basic skills. If students are drilled on limited ranges of subject matter, scores will tend to rise. The more basic the tests are (that is, the easier they are), the more that improvements among the least skilled will affect the mean. Also, the higher the stakes facing a school—and the No Child Left Behind Act makes those stakes very high indeed—the greater will be the incentives for administrators to use some of the many resources at their disposal to make the results come out right, through the judicious manipulation of suspensions and absences, and through outright cheating (yes, it has been known to happen). Some convergence in black and white test scores will probably occur, but partitioning that effect among the competing explanations is a task that will take a few years. Insofar as the convergence has been the result of teaching to the test and of artifacts, it will be temporary.

38 In a given year, IQ tests and academic tests administered to the same sample will produce similar results. Thus, it is possible to make a reasonably good guess about a person's IQ based on his SAT score compared to the distribution of SAT scores in a given year, and after taking the composition of the SAT population into account. But the results of academic tests are sensitive to changes in academic achievement, whereas IQ tests are explicitly designed to measure a general mental factor, *g*, that is independent of academic achievement. A notorious illustration of the way that academic test scores can drop is the period during the 1960's and 1970's when SAT scores declined substantially, even after accounting for changes in the pool of test-takers (Murray and Herrnstein 1992). The intelligence of American youth was not declining, just their academic achievement.

39 The significance of *g*-loadings is discussed later in the text. In terms of

interpreting trends over time, the problem is that tests are not equally good measures of *g*. They go from poor (e.g., a basic reading test) to excellent (the most highly *g*-loaded, individually administered IQ test). It is as if you were trying to measure changes in average height with measuring tapes of varying accuracy. For a statement of the no-change position, see Gottfredson (2005a), or a summary of her argument in Gottfredson (2005b).

40 The .97 figure comes from my analysis of the proxy AFQT score in the most recent release of the 1997 cohort of the National Longitudinal Study of Youth (NLSY). I call it a proxy score because, eight years after the test battery was administered, the Armed Forces still has not gotten around to creating an official AFQT score. The version created by the NLSY staff is a composite of the same subtests used for previous versions of the AFQT, and takes the subject's age into account. The NLSY has released the percentile scores, which I converted to standard scores. The analysis used the NLSY's sample weights to make the results representative of the national population. The NLSY data can be downloaded online.

41 I take the 1.10 figure from Roth, Bevier, Bobko *et al.* (2001), a meta-analysis of the black-white difference in both achievement tests and IQ tests. The Roth *et al.* results are necessarily reflective of pencil-and-paper tests, because that is where the overwhelming majority of published test data come from. With rare exceptions, the data on individually administered IQ tests such as the Wechsler, Stanford-Binet, and Woodcock-Johnson are limited to their periodic standardization samples. The number of such studies is small. These results are overwhelmed in a meta-analysis by the many more studies based on pencil-and-paper tests.

The previous re-norming of the AFQT occurred in 1979, when the AFQT was administered to the 1979 cohort of the NLSY. Herrnstein and Murray (1994) put the black-white difference for that cohort at 1.21 standard deviations. Compared with that figure, the improvement in the 1997 cohort (a .97 black-white difference) is .24 standard deviations. But Neal (in press) has uncovered patterns in the answers of black members of the 1979 cohort that indicate the 1979 cohort produced an artificially low black mean.

First, some background: Any test that tries to measure cognitive ability has to make assumptions about baseline skills. If a person can read, even if not very well, then an IQ test can make use of written items; if the subject is illiterate, it cannot. Similarly, if a person knows numbers and the principles of basic arithmetic, even if not very well, then an IQ test can make use of numeric problems; but if the subject is innumerate, it cannot.

Neal argues that the pattern of answers for the 1979 cohort indicates that "a substantial fraction of the NLSY79 sample of black males who took the ASVAB test lacked the basic math and reading skills covered by the exam, lacked any motivation to put forth effort during the exam, or both," with a similar situation, not quite as bad, for black females (Neal, in press: 13) Given the convergence in academic test scores during the 1980's, it is likely that the proportion of the 1997 NLSY cohort so completely lacking in the basic skills was smaller than in the 1979 cohort. If so, this change alone, not an increase in cognitive ability, would produce convergence in the black-white difference in the AFQT. In addition, the administration of the ASVAB in 1997 was computer-adaptive. Instead of being confronted with pages of questions (105 of them) as in the traditional paper-and-pencil ASVAB (the kind used in 1979), subjects saw one question at time, and the difficulty of each subsequent question was adapted to the subject's previous answer—a method less likely to provoke the kind of give-up response that Neal found in the 1979 data. Neal did not try to estimate the magnitude of the artifact in the 1979 data, but if a "substantial fraction" of the NLSY males had unrealistically low scores, some figure lower than 1.21 standard deviations would be appropriate as a baseline for comparing the 1997 AFQT results. The overall black-white difference of 1.10 standard deviations as found in the meta-analysis is the natural choice.

42 The black and white means on the WISC-IV's measure of full-scale IQ were 91.7 and 103.2 respectively (Prifitera, Weiss, Saklofske *et al.* 2005: 24). Standard deviations for computing the black-white difference were supplied by the Psychological Corporation, which produces the Wechsler tests.

43 The 1.05 and .99 figures come from my analysis of data for the 2001 standardization sample for the Woodcock-Johnson III (WJ-III) test of cognitive ability, provided courtesy of the Woodcock-Munoz Foundation. The results from the WJ-III are noteworthy because the WJ-III provides the best known statistical estimate of *g*. Uniquely among the major standardized tests, the scoring system for the WJ-III uses principal-components analysis to find the best weighted combination of subtests instead of treating all subtests equally (Schrank, McGrew, and Woodcock 2001).

44 Two resourceful defenders of the environmental hypothesis about the black-white difference, James Flynn and William Dickens, are working on their own analysis of the black-white difference over time that should materially add to the state of knowledge when it is released. Here are a few examples of the ambiguities that complicate the assessment of whether the IQ difference has changed, and that have prevented me from stating a confident conclusion:

Example 1. One of the few sources that has several data points over time

with a consistent measure is the General Social Survey (GSS) available online, conducted annually by the National Opinion Research Center; which in most years through the 2000 survey, it included a ten-item vocabulary test.

Example 2. The Kaufman Assessment Battery for Children (K-ABC) is a test that has consistently shown smaller black-white differences than other IQ tests. There are a number of reasons for this, one being that subtests showing large black-white differences were excluded (the K-ABC includes forward-digit span but not backward-digit span, for example). See Jensen (1984) for a full discussion. But though the black-white difference is smaller, it has not changed. In the manual for the original standardization published in 1983, the means on the "Mental Processing Composite" (K-ABC's version of an IQ score) for the white and black samples were 102.0 and 95.0 respectively (A. S. Kaufman and Kaufman 1983: 152). Twenty-one years later, those means were both within a point of their 1983 values—102.7 and 94.8 respectively (A.S. Kaufman and N.L. Kaufman 2004: 96). Which is more meaningful? The smaller black-white difference shown by the K-ABC? Or the absence of any convergence over time?

Example 3. In trying to discriminate between increases in IQ and improvements in academic achievement, one strategy is to explore which parts of the distribution of scores show the most change. Convergence that occurs because of improvements at the bottom of the distribution is likely to reflect remediation of fundamental educational deficits, which could leave the IQ distribution more or less untouched.

In their analysis of six major cross-sectional databases spanning the period from 1965 to 1992, Hedges and Nowell (1998) found that "Racial disparities have diminished over time in the lower tail, but not in the upper tail" (159). In the NAEP, they found that "From 1980 to 1988 there was a substantial increase at all points on the black distribution, with much greater change in the lower percentiles" for the reading scores, and a similar pattern for math scores (161).

Another analysis, however, finds that almost all of the improvement in scores has occurred among black students in the upper half of the black distribution. For example, the AFQT math score of a black male age 15–17 at the 70th percentile of the black distribution in 1980 was equivalent to the score of a white male at about the 28th percentile of the white distribution (Neal, in press, Figure 2a). In 1997, a black male at the 70th percentile of the black distribution had risen to about the 40th percentile of the white distribution. Neal finds a similar result for math scores in the NAEP in the period 1978–1992/96 (Figures 2c and 2d). In contrast, Neal has found almost no increases among students in the bottom half of the black distribution.

How can the results from two analyses be so different? The apparent contradiction—it is not a real contradiction—arises from the fact that almost all of the improvement of blacks in the upper half of the black distribution represents improvement in scores in the lower half of the national distribution of scores. But return to the example of the AFQT: even in 1997, a black subject with a score that put him at the 50th percentile of the white distribution—in other words, a little above the overall national mean—was at about the 80th percentile of the black distribution. In 1980, a black student had to be at about the 90th percentile of the black distribution to have a score above the national mean.

Which analysis should one use? That depends on the topic for which one wants information. If the question is, "Who improved their scores relative to whites, the students at the bottom of the black distribution or the students at the top of it?," Neal's analysis provides the correct answer. If the question is, "Did most of the improvement in black scores occur at the bottom or the top of the *national* distribution of scores?," then Hedges and Nowell's approach provides the correct answer.

In deciding whether IQ has risen, how does one balance these results? I am an optimist about the recent past. To me, the various ambiguous indicators add up to the likelihood that a reduction in the IQ gap has occurred alongside the reduction in the academic-achievement gap. Forced to make a bet, I would guess that the black-white difference in IQ has dropped by somewhere in the range of .10–.20 standard deviations over the last few decades. I must admit, however, that I am influenced by a gut-level conviction that the radical improvement in the political, legal, and economic environment for blacks in the last half of the 20th century must have had an effect on IQ. To conclude that no narrowing whatsoever has occurred raises the question, "How can that be?" One would have to argue that all of the gains in some aspects of the environment have been counterbalanced by new deficits in other aspects, and that those new deficits affect different socioeconomic classes similarly. If the argument is restricted to environmental changes, I cannot imagine how that case might be made.

Another possibility is that improvement in the environmental causes of IQ has been counterbalanced by what is known as "dysgenic" fertility. For several decades at least, women with the highest IQs have been having the fewest babies, and black women have been no different from anyone else (Herrnstein and Murray 1994: chapter 15). But the problem is especially acute among blacks because it is not just black women above the national average IQ who are having the fewest babies but women above the black average. Consider the results for the women of the 1979 NLSY cohort, whose childbearing years are effectively over (they ranged in age from thirty-eight to forty-five when these numbers were collected). Using a nationally representative subsample for the analysis, one finds

that the mean AFQT score of the black women was 85.7. Sixty percent of the children born to this cohort were born to women with AFQT scores below that average. Another 33 percent were born to women with scores from 85.7 to 100. Only 7 percent were born to women with IQs of 100 and over.

Did the children do better? A total of 716 of them were tested with a highly *g*-loaded verbal test, the Peabody Picture Vocabulary Test (revised). The mean of the subset of mothers whose children were tested was 83.7. The mean of their children was 80.2. The mothers and children were tested with different instruments, so it should not be concluded that the black mean actually went down in the new generation. But these data certainly give no reason to think it went up.

It is thus technically possible that black IQ could have remained about the same during the last half-century despite the revolutionary changes for the better in the status of black Americans. Deciding whether that in fact happened requires more evidence than I have presented here.

When I try to forecast the future, I become a pessimist. Here is how I read the overall patterns of change in the academic achievement tests versus the IQ tests:

In a world where Rushton and Jensen are right and the black-white difference is 50- to 80-percent genetic, academic performance and IQ will both improve as the environment improves, and for the same reason: environment plays a role in both measures. Academic test scores will begin to rise before IQ does, because academic performance can improve immediately upon getting a better education whereas the environmental factors affecting IQ are more diffuse. For a related reason—changes in the quality of education can cause substantial increases or drops in academic achievement, whereas IQ cannot be changed much by any known discrete, time-limited environmental change—convergence will be greater in academic achievement than in IQ. Since the environmental role is only 20 to 50 percent of the total, the improvements in both academic and IQ test scores will eventually level off as the limits of environmental change are reached.

To me, the pattern we have observed since good longitudinal data became available in the early 1970's is consistent with these expectations. The only surprise is that evidence for convergence in IQ scores has been so slow to emerge and so spotty. I interpret the pattern as indicating that convergence is nearing an asymptote and that not much will change in the future.

45 Blacks and whites have different distributions of socioeconomic status (SES), and SES is correlated with IQ among both blacks and whites.

When the difference in black and white SES distributions is statistically controlled, studies have typically found that the black-white difference is reduced by about a third of a standard deviation. But when blacks and whites of similar socioeconomic status are compared with each other, the difference as measured in standard deviations remains the same or increases as SES goes up. For a review of the evidence on this point, see Herrnstein and Murray (1994): 286–89.

46 I put aside here the explanation that has received the most publicity in recent years, the phenomenon labeled "stereotype threat." Its discoverers, Claude Steele and Joshua Aronson, demonstrated experimentally that test performance by academically talented blacks was worse when a test was called an IQ test than when it was innocuously described as a research tool (Steele and Aronson 1995). Press reports erroneously interpreted this as meaning that stereotype threat explained away the black-white difference. In reality, Steele and Aronson showed only that it increases the usual black-white difference; if one eliminates stereotype threat, the usual difference remains.

The misrepresentation of these results in the mainstream media was grotesque. For example, the narrator of the PBS television program *Frontline* told his viewers that "blacks who believed the test was merely a research tool did the same as whites." The *Boston Globe* reported that "Black students who think a test is unimportant match their white counterparts' scores." *Newsweek* reported that "blacks who were told that the test was a laboratory problem-solving task that was not diagnostic of ability scored about the same as whites." Such claims have now infiltrated major psychology texts. The third edition of *Psychology* by Davis and Palladino (2002) reports that "The results revealed that African-American students who thought they were simply solving problems performed as well as white students." Similar statements have appeared in scientific journals. All of the above examples are taken from Sackett, Hardison, and Cullen (2004). Sackett *et al.* also have a nice description of how the research results should have been described: "In the sample studied, there are no differences between groups in prior SAT scores, as a result of the statistical adjustment. Creating stereotype threat produces a difference in scores; eliminating threat returns to the baseline condition of no difference" (9).

Readers may follow the latest in the debate by reading a set of responses to Sackett, Hardison, and Cullen (2004) in the April 2005 issue of *American Psychologist*, but nothing in the critiques overturns the above description. The existence of stereotype threat has indeed been demonstrated. It is an interesting phenomenon, and some claims have been made that reducing stereotype threat can improve scores on certain tests (Good, Aronson, and Inzlicht 2003), but the widespread assertion that stereotype threat explains a significant part of the observed black-white difference is wrong. The dissemination of that false assertion is

perhaps understandable in the case of journalists who are not supposed to be sophisticated about such topics. It is less easily explained away when done by authors of technical articles and textbooks.

47 Ogbu (2003).

48 Sowell (2005).

49 Neisser, Boodoo, Bouchard *et al.* (1996): 95.

50 Neisser, Boodoo, Bouchard *et al.* (1996): 95. In truth, the closest thing to direct evidence involves brain size, which is known to have a correlation with IQ (see note 12) and to be different for blacks, whites, and East Asians. See J.P. Rushton and E.W. Rushton (2003) for a recent literature review of the evidence. But the task force did not mention brain size. There is also no mention of IQ in sub-Saharan Africa, the results of transracial adoption studies, the correlation of the black-white difference with the *g*-loadedness of tests, regression to racial means across the range of IQ, or other relevant data. What the task force chose to define as "direct evidence" was a study of children of American black soldiers born to German women after World War II, and studies that use blood-group methods to estimate the degree of African ancestry in American blacks. Both are discussed at length in Rushton and Jensen (2005a) and Nisbett (2005).

51 Rushton and Jensen (2005a).

52 The other articles are Sternberg (2005), Nisbett (2005), Suzuki and Aronson (2005), Gottfredson (2005b), and Rushton and Jensen (2005b)

53 The ten categories, following Rushton and Jensen's wording, are as follow: (1) the world-wide evidence of a consistent black-white-Asian difference, (2) the greater black-white difference on *g*-loaded subtests than on culture-bound subtests, (3) the greater black-white difference on highly heritable subtests than on culturally malleable subtests, (4) the association of the black-white-Asian difference with differences in brain size, (5) the persistence of the black-white-Asian difference among trans-racial adoptees, (6) the consistency of the black-white difference with studies of racial admixture, (7) regression of black and white relatives (offspring or siblings) to their respective racial means, (8) consistency of the black-white-Asian IQ differences with differences in 60 other behavioral traits, (9) consistency of the black-white-Asian differences with evolutionary explanations, and (10) the inability to explain black-white-Asian differences with a zero-genetic model or even with a 50-percent environmental model.

54 Rushton has posted all of the articles at his website.

55 Chakraborty, Kamboh, Nwankwo *et al.* (1992), Parra, Marcini, Akey *et al.* (1998).

56 A variety of studies, summarized in Rushton and Jensen (2005a): 260–61, generally show that the IQs for mixed-race children are about midway between those of children with two white and two black parents. On the other hand, studies that characterized racial composition based on blood group do not predict IQ (Nisbett 2005: 306–07).

57 The results of such a study would be especially powerful if the study also characterized variables like skin color, making it possible to compare the results for subjects for whom genetic heritage and appearance are discrepant. For example, suppose it were found that light-skinned blacks do better in IQ tests than dark-skinned blacks even when their degree of African genetic heritage is the same. This would constitute convincing evidence that social constructions about race, not the genetics of race, influence the development of IQ. Given a well-designed study, many such hypotheses about the conflation of social and biological effects could be examined.

58 Spearman (1927): 379.

59 The average adult gets a digits-backward score of 5 (Jensen 1998: 263). You may compare your own score with the highest I have observed, 13 and 12, achieved respectively by José Zalaquett, former chairman of Amnesty International, and the political analyst Charles Krauthammer. Zalaquett's score might have been higher if he had not been in a car weaving through traffic at 70 miles per hour on the New Jersey Turnpike. Krauthammer's score might have been higher if he hadn't been driving.

60 Jensen (1998): 370.

61 A similarly clean example of a black-white difference is produced by reaction-time tests, in which two different measures are taken: the time it takes for the subject to respond to the lighted buttons that constitute the stimulus (a *g*-loaded measure) and the time it takes to move one's finger from the home button to the appropriate lighted button (no *g*-loading). Black subjects have faster movement times and slower response times—once again a contrast, consistent with Spearman's hypothesis, produced at the same time with the same examiner in the same setting. None of the usual ways to explain away the black-white difference through cultural causes applies. See Jensen (1998): 389–93.

62 Jensen (1998): 369–402.

63 Nyborg and Jensen (2000). It should also be noted that one test of Spearman's hypothesis has been conducted comparing East Asians and

whites. The better the measure of *g*, the greater the advantage of East Asians over whites. See Nagoshi, Johnson, DeFries *et al.* (1984).

64 Jensen's evidence has been accompanied by a debate over his method of correlated vectors for testing Spearman's hypothesis. P.H. Schönemann has argued, most extensively in Schönemann (1997), that Jensen's evidence was no more than a statistical artifact, a claim refuted by Dolan and Lubke (2001). But other ways in which the method of correlated vectors might yield spurious results are still being debated; e.g., Dolan (2000), Lubke, Dolan, and Kelderman (2001), Dolan, Roorda, and Wicherts (2004), Ashton and Lee (in press). These arguments are being carried on at an arcane methodological level. I am making a limited claim about what Jensen has established beyond dispute: when you take a battery of mental tests, subject them to a factor analysis, and correlate the loadings on the first factor with the size of the black-white difference, the correlation will average about .6. The actual method of correlated vectors is more complicated than this, and is described in Jensen (1998): 372–74.

65 Factor analysis can be conducted in many different ways, which has led to widespread popular acceptance of one of Stephen Jay Gould's allegations in his best-selling book, *The Mismeasure of Man* (1981), namely, that *g* is a statistical artifact that appears only when certain analytic choices are made. Actually, the opposite is true. A single factor, typically explaining about three times as much variance as all the other factors combined, emerges under all of the normal methods of conducting a factor analysis. The only exception occurs if the factor-analysis program is explicitly instructed to apportion the variance in such a way that a single factor does not emerge. But if you do that and then try to publish your results, the reviewers will point out that if you hadn't issued that instruction, you would have gotten a dominant single factor. As Richard Herrnstein liked to say, "You can make *g* hide, but you can't make it go away." For a review of this issue with sources, see the Afterword to the softcover edition of *The Bell Curve* (559–62). For a technical demonstration of the convergent results from alternative ways of conducting a factor analysis, see Ree and Earles (1991). For a wide-ranging set of articles about the current role of *g* in understanding intelligence, see the articles in the special section of the January 2004 issue of *Journal of Personality & Social Psychology* commemorating the 100th anniversary of Spearman's discovery of *g*. An overview is given in Lubinski (2004).

66 Gould (1981) still shapes the lay received wisdom about IQ tests, but his denunciation of *g* was already technically outdated when it was published. For an account of the differing ways in which *The Mismeasure of Man* was assessed by the media and by scholars, see Davis (1983). For a recent discussion of the nature of *g* and the issues that Gould was wrong about, see Bartholomew (2004).

67 Jensen (1998): 182–89.

68 Jensen (1998): 137–68.

69 Haier, Jung, Yeo *et al.* (2004); Thompson, Cannon, Narr *et al.* (2001).

70 Let it be clear: I am *not* asserting that putting these two facts together proves that the black-white difference is genetic. The logic of the situation was memorably converted to an analogy in Lewontin (1970) and adapted in Herrnstein & Murray (1994): 298. If you take two handfuls of genetically identical seed corn and plant one in Iowa and the other in the Mohave Desert, you will get a large group difference in results despite the high heritability of the traits of corn. William Dickens and James Flynn have operationalized the analogy through a simulation model that produces a large black-white difference from environmental factors even given high heritability (Dickens and Flynn 2001). The validity of that model was subsequently disputed by Loehlin (2002) and Rowe and Rodgers (2002), with a reply by Dickens and Flynn (2002). But that debate does not pertain here. The implications I describe follow simply from knowing that *g* is highly heritable among blacks, as it is among all groups, and that the black-white difference is largely a difference in *g*.

71 See te Nijenhuis, Voskuijl, and Schijve (2001), who also found evidence, as did Neubauer and Freudenthaler (1994), that coaching also reduced the *g*-loadedness of the test, and for the obvious reason: noise has been introduced into the IQ score, changing the score but not the thing that makes an IQ test predictive, *g*. An athletic analogy may be usefully pursued for understanding these results. Suppose you have a friend who is a much better athlete than you, possessing better depth perception, hand-eye coordination, strength, and agility. Both of you try high-jumping for the first time, and your friend beats you. You practice for two weeks; your friend doesn't. You have another contest and you beat your friend. But if tomorrow you were both to go out together and try tennis for the first time, your friend would beat you, just as your friend would beat you in high-jumping if he practiced as much as you did.

72 Flynn (1984) is an early statement. Over the years since *The Bell Curve* was published, it has been especially exasperating to be told, or to see it written, that Herrnstein and I were wrong because we did not know about the Flynn effect. We not only provided the first discussion of the Flynn effect aimed at a general audience; we *named* it (Herrnstein and Murray 1994: 307–09). Some scholars, notably J. Philippe Rushton, have subsequently called it the "Lynn-Flynn effect," thereby acknowledging Richard Lynn's role in identifying the rise in IQ scores.

73 Flynn (1998).

74 An early statement of this evidence, based on analysis of the *g* loadings of subtests, is Jensen (1998): 320–21. Rushton (1999) elaborates, disputed in Flynn (1999) and Flynn (2000), with a rejoinder in Rushton (2000). Since then the evidence that the Flynn effect does not consist of increases in *g* has been augmented by an independent method, multigroup confirmatory factor analysis (MGCFA), which permits a test for factorial invariance between cohorts. In less technical terms, the method tests for whether differences in IQ scores between groups reflects true differences in *g.* See Lubke, Dolan, Kelderman *et al.* (2003) for a description of the method and its uses. Wicherts, Dolan, Hessen *et al.* (2004) used MGCFA on five large databases: Dutch adults in 1967/68 and 1998/99; Danish draftees in 1988 and 1998; Dutch high-school students in 1984 and 1994/95; Dutch children in 1981/82 and 1992/93; and Estonian children 1934/36 and 1997/98. The authors found that the hypothesis of factor invariance was untenable, and that the gains in intelligence-test scores were not manifestations of increases in *g*. Previously, Dolan (2000) and Dolan & Hamaker (2001) had used the MGCFA to test for factor invariance between blacks and whites on IQ tests, and had concluded that the results passed the MGCFA test. In other words, the black-white differences were consistent with a difference in *g*. It was this contrast in results that led Wicherts and his colleagues to conclude that the Flynn effect would have little effect on the black-white difference.

75 Wicherts, Dolan, Hessen *et al.* (2004): 531.

76 In the text I ignore Europe, where both academic and political elites have suppressed the discussion of group differences even more effectively than in America. Contemporaneously, the European Union has revolutionized free movement within Europe. That, combined with immigration from outside Europe, legal and illegal, has produced unprecedented population change in countries that historically have been ethnically homogeneous.

Immigration poses problems for European countries that are qualitatively different from those faced by the United States. Becoming an American requires only that immigrants buy into a set of American ideals. You can move to America from anywhere in the world, be of any ethnicity, social class, or race, and become an American. Assimilation is what America does—not as well as it used to, but still pretty well. The European Union's immigration policy has, willy-nilly, decided that now you can move to Denmark and become Danish or move to France and become French. Is this true? Everyday experience suggests that Denmark's culture works because it fits the characteristics of Danes, that France's culture works because it fits the characteristics of the French, and that these ethnic characteristics are importantly different and deeply rooted, whether in genes or in habits of the heart. Replace a large proportion of French with Danes—let alone peoples more distant—and French culture

will be profoundly changed. But it is taboo among the elites to talk about such things (although ordinary people sense what is at stake), and so a momentous social experiment is under way without any reason to think that its assumptions are correct, many historical reasons for thinking they are wrong, and recurring stories on the evening news suggesting that the social fabrics of Europe will be shredded before the elites can make themselves come to grips with what they have been doing.

77 A few systematic examinations of this issue have been published; e.g., Lott (2000) on the effects of affirmative action on policing. For a journalistic account of the effects of political correctness on the Los Angeles Police Department, see Golab (2005).

78 Sommers (2001).

79 Satel (2002).

80 For examples of the effects of controlling for group differences on a variety of outcomes and groups, see Herrnstein and Murray (1994): chapter 14, Nyborg and Jensen (2001), and Kanazawa (2005).

81 See Pinker (2002): chapter 16 for a discussion of how politics interacts with the acceptance of group differences.

82 Pinker (2002): 340.

Bibliography

Ankney, D. 1992. "Sex differences in relative brain size: The mismeasure of woman, too?" *Intelligence* 16: 329–36.

Ashton, M.C. and K. Lee. in press. "Problems with the method of correlated vectors." *Intelligence*.

Bamshad, M.J., S. Wooding, W.S. Watkins, *et al.* 2003. "Human population genetic structure and inference of group membership." *American Journal of Human Genetics* 72: 578–89.

Baron-Cohen, S. 2003. *The Essential Difference: Male and Female Brains and the Truth about Autism*. New York: Basic Books.

Bartholomew, D.J. 2004. *Measuring Intelligence: Facts and Fallacies*. Cambridge: Cambridge Univ. Press.

Benbow, C.P., D. Lubinski, D.L. Shea, *et al.* 2000. "Sex differences in mathematical reasoning ability at age 13: Their status 20 years later."

Psychological Science 11: 474-80.

Bowcock, A.M., A. Ruiz-Linares, J. Romfohrde*, et al.* 1994. "High resolution of human evolutionary trees with polymorphic microsatellites." *Nature* 368: 455–57.

Brown, R.P. and R.A. Josephs. 1999. "A burden of proof: Stereotype relevance and gender differences in math performance." *Journal of Personality & Social Psychology* 76: 246–57.

Byrnes, J.P., D.C. Miller, and W.D. Schafer. 1999. "Gender differences in risk taking: A meta-analysis." *Psychological Bulletin* 125: 367–83.

Cahill, L. 2005. "His brain, her brain." *Scientific American*, May.

Calafell, F., A. Shuster, W.C. Speed*, et al.* 1998. "Short tandem repeat polymorphism evolution in humans." *European Journal of Human Genetics* 6: 38–49.

Casey, M.B., R. Nuttall, E. Pezaris*, et al.* 1995. "The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples." *Developmental Psychology* 31: 697–705.

Chakraborty, R., M.I. Kamboh, M. Nwankwo*, et al.* 1992. "Caucasian genes in American blacks." *American Journal of Human Genetics* 50: 145–55.

Dabbs, J.M. and M.G. Dabbs. 2000. *Heroes, Rogues, and Lovers: Testosterone and Behavior*. New York: McGraw-Hill.

Davis, B.D. 1983. "Neo-Lysenkoism, IQ, and the Press." *The Public Interest* no. 73: 41–59.

Dickens, W.T. and J.R. Flynn. 2001. "Heritability estimates versus large environmental effects: The IQ paradox resolved." *Psychological Review* 108: 346-69.

—. 2002. "The IQ paradox is still resolved: Reply to Loehlin (2002) and Rowe and Rodgers (2002)." *Psychological Review* 109: 764-71.

Dolan, C.V. 2000. "Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis." *Multivariate Behavioral Research* 35: 21–50.

Dolan, C.V. and G.H. Lubke. 2001. "Viewing Spearman's hypothesis from the perspective of multigroup PCA: A comment on Schoenemann's criticism." *Intelligence* 29: 231–45.

Dolan, C.V., W. Roorda, and J.M. Wicherts. 2004. "Two failures of Spearman's hypothesis: The GATB in Holland and the JAT in South Africa." *Intelligence* 32: 231–45.

Edwards, A.W.F. 2003. "Human genetic diversity: Lewontin's fallacy." *BioEssays* 25: 798–801.

Flynn, J.R. 1984. "The mean IQ of Americans: Massive gains 1932 to 1978." *Psychological Bulletin* 95: 29-51.

—. 1998. "IQ gains over time: Toward finding the causes" in *The Rising Curve: Long-term Gains in IQ and Related Measures*, edited by U. Neisser. Washington: American Psychological Association.

—. 1999. "Evidence against Rushton: The genetic loading of WISC-R subtests and the causes of between-group IQ differences." *Personality and Individual Differences* 26: 373–79.

—. 2000. "IQ gains and fluid *g*." *American Psychologist* 55: 543.

Friedman, L. 1995. "The space factor in mathematics: Gender differences." *Review of Educational Research* 65: 22–50.

Fryer, R.G. and S.D. Levitt. 2004. "Understanding the black white test score gap in the first two years of school." *Review of Economics and Statistics* 86: 447–64.

Gallagher, A.M. and J.C. Kaufman (eds). 2005. *Gender Differences in Mathematics: An Integrative Psychological Approach.* Cambridge: Cambridge Univ. Press.

Geary, D.C. 1998. *Male, Female: The Evolution of Human Sex Differences*. Washington: American Psychological Association.

Geary, D.C., S.J. Saults, F. Liu, *et al.* 2000. "Sex differences in spatial cognition, computational fluency, and arithmetical reasoning." *Journal of Experimental Child Psychology* 77: 337–53.

Golab, J. 2005. "How racial PC corrupted the LAPD." *American Enterprise*. June.

Goldstein, J.M., L.J. Seidman, N.J. Horton, *et al.* 2001. "Normal sexual dimorphism of the adult human brain assessed by *in vivo* magnetic resonance imaging." *Cerebral Cortex* 11: 490–97.

Good, C., J. Aronson, and M. Inzlicht. 2003. "Improving adolescents' standardized test performance: An intervention to reduce the effects of

stereotype threat." *Journal of Applied Developmental Psychology* 24: 645–62.

Gottfredson, L.S. 2005a. "Implications of cognitive differences for schooling within diverse societies" in *Comprehensive Handbook of Multicultural School Psychology*, edited by C. L. Frisby and C. R. Reynolds. New York: Wiley.

—. 2005b. "What if the hereditarian hypothesis is true?" *Psychology, Public Policy, and Law* 11: 311–19.

Gould, S.J. 1981. *The Mismeasure of Man*. New York: W. W. Norton.

Gron, G., M. Spitzer, R. Tomczak, *et al.* 2000. "Brain activation during human navigation: Gender-different neural networks as a substrate of performance." *Nature Neuroscience* 3: 404–08.

Haier, R.J., R.E. Jung, R.A. Yeo, *et al.* 2004. "Structural brain variation and general intelligence." *NeuroImage* 23: 425–33.

Halpern, D.F. 2000. *Sex Differences in Cognitive Ability*. Mahwah, NJ: L. Erlbaum Associates.

Hedges, L.V. and A. Nowell. 1998. "Black-White Test Score Convergence since 1965." Pp. 149–81 in *The Black-White Test Score Gap*, edited by C. Jencks and M. Phillips. Washington: Brookings Institution Press.

Herrnstein, R.J. and C. Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.

Hrdy, S.B. 1999. *Mother Nature: A History of Mothers, Infants, and Natural Selection*. New York: Pantheon Books.

Jensen, A.R. 1984. "The Black-White difference on the K-ABC: Implications for future tests." *Journal of Special Education* 18: 377-408.

—. 1998. *The* g *Factor: The Science of Mental Ability*. Westport, CT: Praeger.

Johnson, E.S. 1984. "Sex differences in problem solving." *Journal of Educational Psychology* 76: 1359–71.

Jones, C.M., V.A. Braithwaite, and S.D. Healy. 2003. "The evolution of sex differences in spatial ability." *Behavioral Neuroscience* 117: 403–11.

Kanazawa, S. 2005. "Is 'discrimination' necessary to explain the sex gap

in earnings?" *Journal of Economic Psychology* 26: 269–87.

Kaufman, A.S. and N.L. Kaufman. 1983. *Kaufman Assessment Battery for Children: Interpretive Manual*. Circle Pines, MN: AGS Publishing.

—. 2004. *KABC-II Manual*. Circle Pines, MN: AGS Publishing.

Kimball, M.M. 1989. "A new perspective on women's math achievement." *Psychological Bulletin* 105: 198–214.

Kimura, D. 1999. *Sex and Cognition*. Cambridge, MA: MIT Press.

Leroi, A.M. 2005. "A family tree in every gene." Pp. 21 in *New York Times*, March 14.

Lewontin, R.C. 1970. "Race and intelligence." *Bulletin of the Atomic Scientists* 26: 2-8.

—. 1972. "The apportionment of human diversity." *Evolutionary Biology* 6: 381–98.

Linn, M.C. and A.C. Peterson. 1985. "Emergence and characterization of sex differences in spatial ability: A meta-analysis." *Child Development* 56: 1479–98.

Loehlin, J.C. 2002. "The IQ paradox: Resolved? Still an open question." *Psychological Review* 109: 754–58.

Lott, J.R. 2000. "Does a helping hand put others at risk? Affirmative action, police departments, and crime." *Economic Inquiry* 38: 239–77.

Lubinski, D. 2004. "Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) '"General intelligence," objectively determined and measured'." *Journal of Personality & Social Psychology* 86: 96–111.

Lubinski, D., C.P. Benbow, D.L. Shea*, et al.* 2001. "Men and women at promise for scientific excellence: Similarity not dissimilarity." *Psychological Science* 12: 309-17.

Lubke, G.H., C.V. Dolan, and H. Kelderman. 2001. "Investigating group differences using Spearman's hypothesis: An evaluation of Jensen's method." *Multivariate Behavioral Research* 36: 299–324.

Lubke, G.H., C.V. Dolan, H. Kelderman*, et al.* 2003. "On the relationship between sources of within- and between-group differences and measurement invariance in the common factor model." *Intelligence* 31.

Lynn, R. 1998. "Has the black-white intelligence difference in the United States been narrowing over time?" *Personality and Individual Differences* 25: 999-1002.

—. 1999. "Sex differences in intelligence and brain size: A developmental theory." *Intelligence* 27: 1–12.

Lynn, R. and P. Irwing. 2004. "Sex differences on the progressive matrices: A meta-analysis." *Intelligence* 32: 481–98.

Matarazzo, J.D. 1972. *Wechsler's Measurement and Appraisal of Adult Intelligence*. New York: Oxford Univ. Press.

McDaniel, M.A. 2005. "Big-brained people are smarter: A meta-analysis of the relationship between *in vivo* brain volume and intelligence." *Intelligence* 33: 337–46.

Mountain, J.L. and L.L. Cavalli-Sforza. 1997. "Multilocus genotypes, a tree of individuals, and human evolutionary history." *American Journal of Human Genetics* 61: 705–18.

Murray, C. 2003. *Human Accomplishment: The Pursuit of Excellence in the Arts and Sciences, 800 B.C. to 1950*. New York: HarperCollins.

Murray, C. and R.J. Herrnstein. 1992. "What's really behind the SAT-score decline." *The Public Interest* no. 106: 32-56.

Nagoshi, C.T., R.C. Johnson, J.C. DeFries, *et al.* 1984. "Group differences and first principle-component loadings in the Hawaii family study of cognition: A test of the generality of 'Spearman's hypothesis.'" *Personality and Individual Differences* 5: 751–53.

Neal, D. in press. "Why has black-white skill convergence stopped?" in *Handbook of Economics of Education*, edited by E. Hanushek and F. Welch. New York: Elsevier.

Neisser, U., G. Boodoo, T.J. Bouchard, Jr., *et al.* 1996. "Intelligence: Knowns and unknowns." *American Psychologist* 51: 77-101.

Neubauer, A.C. and H.H. Freudenthaler. 1994. "Reaction time in a sentence-picture verification test and intelligence: Individual strategies and the effects of extended practice." *Intelligence* 19: 193–218.

Nisbett, R.E. 2005. "Heredity, environment, and race differences in IQ: A commentary on Rushton and Jensen (2005)." *Psychology, Public Policy, and Law* 11: 302–10.

Nyborg, H. In press. "Sex-related differences in general intelligence *g*, brain size, and social status." *Personality and Individual Differences*.

Nyborg, H. and A.R. Jensen. 2000. "Black-white differences on various psychometric tests: Spearman's hypothesis tested on American armed services veterans." *Personality and Individual Differences* 28: 593–99.

—. 2001. "Occupation and income related to psychometric *g*." *Intelligence* 29: 45-55.

Ochse, R. 1990. *Before the Gates of Excellence: The Determinants of Creative Genius*. Cambridge: Cambridge Univ. Press.

Ogbu, J. 2003. *Black American Students in an Affluent Suburb: A Study of Academic Disengagement*. Hillsdale, NJ: Lawrence Erlbaum Assoc.

Parra, E.J., A. Marcini, J. Akey, *et al.* 1998. "Estimating African American admixture proportions by use of population specific alleles." *American Journal of Human Genetics* 63: 1839–51.

Pinker, S. 2002. *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking Penguin.

Prifitera, A., L.G. Weiss, D.H. Saklofske, *et al.* 2005. "The WISC-IV in the clinical assessment context." Pp. 3–32 in *WISC-IV Clinical Use and Interpretation: Scientist-Practitioner Perspectives*, edited by A. Prifitera, D. H. Saklofske, and L. G. Weiss.

Ree, M.J. and J.A. Earles. 1991. "The stability of convergent estimates of *g*." *Intelligence* 15: 271-78.

Rhoads, S.E. 2004. *Taking Sex Differences Seriously*. San Francisco: Encounter Books.

Rowe, D.C. and J.L. Rodgers. 2002. "Expanding variance and the case of historical changes in IQ means: A critique of Dickens and Flynn (2001)." *Psychological Review* 109: 759–63.

Rosenberg, N.A., J.K. Pritchard, J.L. Weber, *et al.* 2002. "Genetic structure of human populations." *Science* 298: 2381–85.

Roth, P.L., C.A. Bevier, P. Bobko, *et al.* 2001. "Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis." *Personnel Psychology* 54: 297–330.

Royer, J.M., L.N. Tronsky, Y. Chan, *et al.* 1999. "Math-fact retrieval as the cognitive mechanism underlying gender differences in math test

performance." *Contemporary Educational Psychology* 24: 181–266.

Rushton, J.P. 1999. "Secular gains in IQ not related to the *g* factor and inbreeding depression—unlike Black-White differences: A reply to Flynn." *Personality and Individual Differences* 26: 381–89.

—. 2000. "Flynn effects not genetic and unrelated to race differences." *American Psychologist* 55: 542-43.

Rushton, J.P. and A.R. Jensen. 2005a. "Thirty years of research on race differences in cognitive ability." *Psychology, Public Policy, and Law* 11: 235–94.

—. 2005b. "Wanted: More race realism, less moralistic fallacy." *Psychology, Public Policy, and Law* 11: 328–36.

Rushton, J.P. and E.W. Rushton. 2003. "Brain size, IQ, and racial-group differences: Evidence from musculoskeletal traits." *Intelligence* 31: 139–55.

Sackett, P.R., C.M. Hardison, and M.J. Cullen. 2004. "On interpreting stereotype threat as accounting for African American-white differences on cognitive tests." *American Psychologist* 59: 7–13.

Satel, S. 2002. *PC, M.D.: How Political Correctness is Corrupting Medicine*. New York: Basic Books.

Schrank, F.A., K.S. McGrew, and R.W. Woodcock. 2001. *Technical Abstract* (Woodcock-Johnson III Assessment Service Bulletin No. 2). Itasca, IL: Riverside Publishing.

Schönemann, P.H. 1997. "Famous artifacts: Spearman's hypothesis." *Cahiers de Psychologie Cognitive* 16: 665–94.

Simonton, D.K. 1984. *Genius, Creativity, and Leadership*. Cambridge, MA: Harvard Univ. Press.

—. 1994. *Greatness: Who Makes History and Why*. New York: Guilford Press.

—. 1999. *Origins of Genius: Darwinian Perspectives on Creativity.* Oxford: Oxford Univ. Press.

Sommers, C.H. 2001. *The War Against Boys: How Misguided Feminism Is Harming Our Young Men*. New York: Simon & Schuster.

Sowell, T. 2005. *Black Rednecks and White Liberals*. San Francisco:

Encounter Books.

Spearman, C. 1927. *The Abilities of Man*. New York: Macmillan.

Steele, C.M. and J. Aronson. 1995. "Stereotype threat and the intellectual test performance of African Americans." *Journal of Personality & Social Psychology* 69: 797–811.

Stephens, J.C., J.A. Schneider, D.A. Tanguay*, et al.* 2001. "Haplotype variation and linkage disequilibrium in 313 human genes." *Science* 293: 489–93.

Sternberg, R.J. 2005. "There are no public policy implications: A reply to Rushton and Jensen (2005)." *Psychology, Public Policy, and Law* 11: 295–301.

Stipek, D.J. and J.H. Gralinski. 1991. "Gender differences in children's achievement-related beliefs and emotional responses to success and failure in mathematics." *Journal of Educational Psychology* 83: 361–71.

Suzuki, L. and J. Aronson. 2005. "The cultural malleability of intelligence and its impact on the racial/ethnic hierarchy." *Psychology, Public Policy, and Law* 11: 320–27.

Tang, H., T. Quertermous, B. Rodriguez*, et al.* 2005. "Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies." *American Journal of Human Genetics* 76: 268–75.

te Nijenhuis, J., O.F. Voskuijl, and N.B. Schijve. 2001. "Practice and coaching on IQ tests: Quite a lot of *g*." *International Journal of Selection and Assessment* 9: 302–08.

Thompson, P., T.D. Cannon, K.L. Narr*, et al.* 2001. "Genetic influences on brain structure." *Nature Neuroscience* 4: 1–6.

Valian, V. 1999. *Why So Slow? The Advancement of Women*. Cambridge, MA: MIT Press.

Wicherts, J.M., C.V. Dolan, D.J. Hessen*, et al.* 2004. "Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect." *Intelligence* 32: 509–37.

Wynn, T.G., F.D. Tierson, and C.T. Palmer. 1996. "Evolution of sex differences in spatial cognition." *Yearbook of Physical Anthropology* 39: 11–42.