Searching for Justice

The Discovery of IQ Gains Over Time

James R. Flynn
University of Otago

Humane-egalitarian ideals, whose aims are group justice and reducing environmental inequality and privilege, must be tested against reality, as revealed by psychology and other social sciences. Four issues are addressed: the equation between IQ and intelligence, whether group potential is determined by a group's mean IQ, whether the Black–White IQ gap is genetic, and the meritocratic thesis that genes for IQ will become highly correlated with class. Massive IQ gains over time test the IQ–intelligence equation, reveal groups who achieve far beyond their mean IQs, and falsify prominent arguments for a genetic racial IQ gap. Class IQ trends suggest America is not evolving toward a meritocracy, but a core refutation of that thesis is needed and supplied. Finally, the viability of humane ideals is assessed against a worst-case scenario.

This article is an attempt to make explicit the connecting thread of 20 years of research into group IQ differences. On one level, the various researches recommended one another. Massive IQ gains over time revealed that the present generation has a huge IQ advantage over the previous generation. Yet the IQ advantage did not seem to be accompanied by a corresponding achievement advantage, which suggested the possibility that the mean IQ of Chinese and Japanese Americans might underpredict their achievements. IQ differences between the generations are clearly environmental in origin. Yet heritability of IQ within generations is robust, which suggested that high within-race heritability estimates do not signal a genetic IQ gap between Black and White populations. From this, it is a natural extension to address the meritocracy thesis of *The Bell Curve: Intelligence and Class in American Life* (Herrnstein & Murray, 1994): This thesis is that the heritability of IQ plus social trends render inevitable a society in which good genes for IQ are highly correlated with class.

However, this tidy sequence of congenial concepts omits emotion and motive. The true connecting thread has been a commitment to a humane-egalitarian concept of social justice. Ever since Plato, those so committed have known that splendid ideals are not enough. They must show that their ideals are viable in the light of everything that the best social and biological science tells us about human society and show that they have the courage to face up to the consequences of putting their ideals into practice. Today, it is psychology rather than philosophy that is updating Plato's *Republic*. A dialogue with Arthur Jensen and Charles Murray, thinkers equally committed to social justice, was almost inevitable. It takes us from IQ gains over time, to the IQs and achievements of Chinese and Japanese Americans, to a review of the race and IQ debate, and finally, to a critique of the meritocracy thesis.

IQ Gains and Measuring Intelligence

The discovery of IQ gains over time emerged naturally from the work of many scholars. Several had evidenced gains but tended to attribute them to a particular group or nation or kind of test. Charles Murray called IQ gains the "Flynn effect" because I infered that they were part of a persistent and perhaps universal phenomenon (Herrnstein & Murray, 1994, p. 307). That inference was more a product of accident than perspicacity.

In 1981, I began an analysis of Armed Forces mental tests that revealed that Blacks had been gaining on Whites (Flynn, 1987a, p. 235). This suggested a survey of Wechsler and Stanford–Binet manuals to see whether armed forces tests correlated with genuine IQ tests. There were no such data, but every manual had a table showing that the publisher's new test had a high correlation with the previous version, for example, that the WISC-R (Wechsler Intelligence Scale for Children–Revised) correlated with the older WISC (Wechsler Intelligence Scale for Children). The tables also showed something persistent and surprising. Without exception, whenever the same participants were given both the new test and an older test, they had a higher score on the latter. If they averaged 100 on the WISC-R, normed in 1972, they averaged 108 on the WISC, normed in 1947 and 1948. The only possible explanation was that representative samples of White Americans were setting higher standards of test performance over time. Children who could barely match a 1972 standardization

Editor's note. William C. Howell served as action editor for this article.

Author's note. The original version of this article was presented at the 105th Annual Convention of the American Psychological Association in Chicago, Illinois, in August 1997. I thank Wendy Williams for the invitation to the convention, Steve Ceci for chairing the session, and Dick Neisser for valuable critical comment.

Correspondence concerning this article should be addressed to James R. Flynn, Department of Political Studies, University of Otago, P.O. Box 56, Dunedin, New Zealand. Electronic mail may be sent to jim.flynn@stonebow.otago.ac.nz.

 0003-066X/99/$2.00
Vol. 54, No. 1, 5–20

James R. Flynn

sample and therefore scored 100 easily exceeded the performance of a 1947 sample and therefore scored 108. Clearly from one sample to the other, over a period of 25 years, Americans had gained 8 IQ points.

The evidence of the test manuals was submitted to the *Harvard Educational Review.* The reader and editor found the evidence fragmentary and unconvincing. This prompted an attempt to locate every study in which the same group of participants had taken two or more Wechsler or Stanford–Binet IQ tests. The results were published in *Psychological Bulletin* (Flynn, 1984) and were supplemented a few years later (Flynn, 1987b). Data from 73 studies containing 7,500 participants ages 2 to 48 years showed that between 1932 and 1978, White Americans had gained 14 IQ points. The rate of gain was about 0.30 IQ points per year, roughly uniform over time and similar for all ages. It posed an extraordinary puzzle, because the last two decades of the period coincided with that sad time when American high school students showed a decline on the Scholastic Aptitude Test (SAT). Subsequently, scholars citing elementary school data have argued against the decline, but no one has posited an improvement that comes near to matching the magnitude of the IQ gains. So how could American students be getting so much brighter, as measured by IQ tests, and yet be learning no more in school, as measured by achievement tests? That the quality of schooling had deteriorated that dramatically seemed absurd. Alternatively, that nonschool factors like TV or family environment could both elevate intelligence and undermine learning seemed even more absurd (Flynn, 1984, 1987b, pp. 176–177).

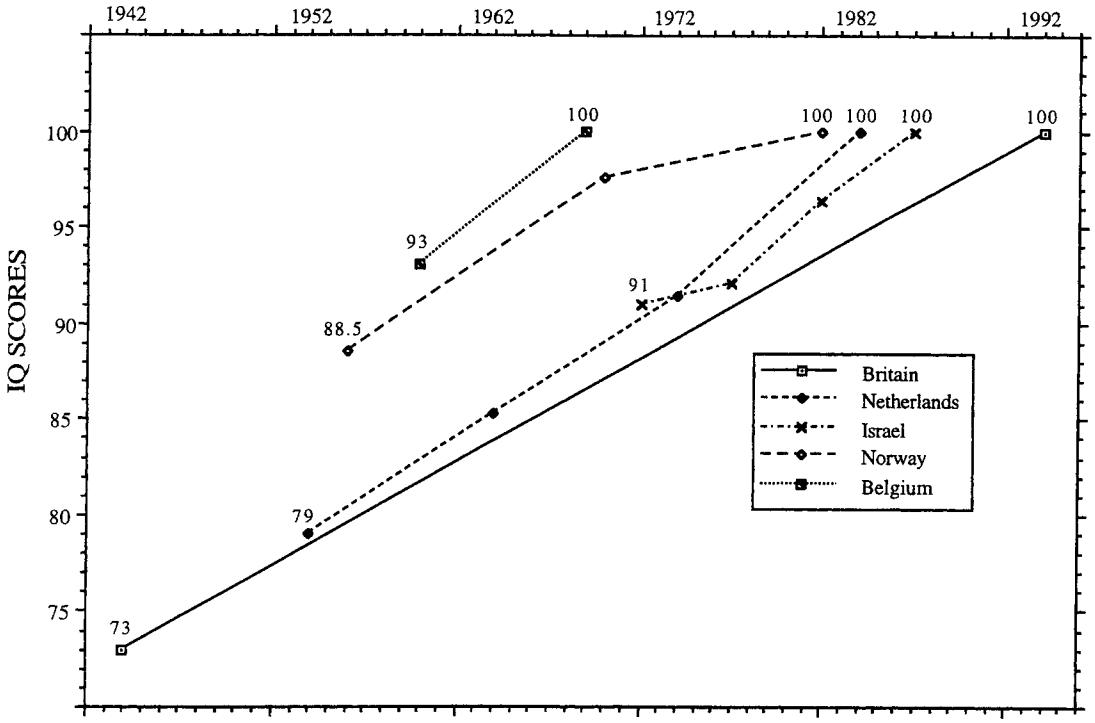
Through personal communications, Arthur Jensen expressed four reservations: (a) The possibility of sample bias should be eliminated by comprehensive samples, such as mass testing of draft registrants; (b) tests should remain unaltered from one generation to another, and estimates of trends should be based on raw score differences; (c) particular emphasis should be placed on using mature participants, participants who have reached the peak of their raw score performance; (d) particular emphasis should be placed on culture-reduced tests like Raven's Progressive Matrices. Jensen's third point reflected a suspicion that the current generation was merely reaching peak performance at an earlier age than the last generation and that at full maturity, the two generations would score about the same. His fourth point reflected his view that tests like Ravens are the purest measure of *g,* the general intelligence factor, and would show no gains over time. Wechsler and Stanford–Binet tests measure intelligence partially through the vehicle of items taught in school, and therefore, more or better schooling could produce the appearance of intelligence gains over time (Flynn, 1987b, p. 171).

In response, data relevant to IQ trends over time were solicited from every nation that had either military mental testing or testing organizations. It was anticipated that this would satisfy Jensen on all but his fourth point, and this proved correct. However, I too assumed that culture-reduced tests would show negligible gains, thanks to their resistance to whatever environmental factors were influencing Wechsler and Stanford–Binet tests. The data from 14 nations, published in *Psychological Bulletin* (Flynn, 1987b) and later expanded to cover 20 nations (Flynn, 1994b), staggered both of us: They showed that from 1930 to the present, the largest IQ gains were on culture-reduced tests like Ravens. The best Ravens data come from four nations. Military samples from the Netherlands, Israel, Norway, and Belgium virtually exhaust cohorts of young adults and reveal what was happening from 1952 to 1982.

As Figure 1 shows, all but Norway enjoyed gains ranging from 18 to 20 IQ points in one generation; Norway was similar until 1968 but after that gained at a generational rate of 7.5 points. However, if you merge the data for all four nations, something odd happens: You get a monotonic rate of gain over the whole 30 years. It is as if some unseen hand is propelling scores upward at a rate of about 6 IQ points per decade, with individual nations scattering randomly around that value. Figure 1 also includes British data which, although of much lower quality, gain credibility from their match for the best data. The British data are of special interest because they cover a period of 50 years and include participants ranging from young adults to adults in old age. John Raven took the test scores of all these participants, those ages 25 to 65 tested in 1942 and those of the same ages tested in 1992, and plotted them by birth date. This gave him scores for those born all the way from 1877 to the 1970s. Raven has defended himself against the criticism that the 1992 sample suffers from score inflation and aduces strong evidence that the bottom half of the Ravens curve was unaffected (Raven, 1995). Therefore, the fifth percentile from 1992 was used to compare the two standardization samples.

Figure 2 shows that the bottom 90% of Britons born in 1877 fall below the fifth percentile of those born in 1967, which is to say below an IQ of 75 calculated on 1967

Figure 1
Five Nations and Matrices Tests: Rates of IQ Gain Compared



Note. Every nation is normed on its own samples. Therefore, although nations can be roughly compared in terms of different rates of IQ gain, they cannot be compared in terms of IQ scores. That is, the fact that the mean IQ of one nation appears higher than another at a given time is purely an artifact. Sources for these data are Flynn (1987b, pp. 172–174), Flynn (1998d), Raven, Raven, and Court (1998, Graph G2).

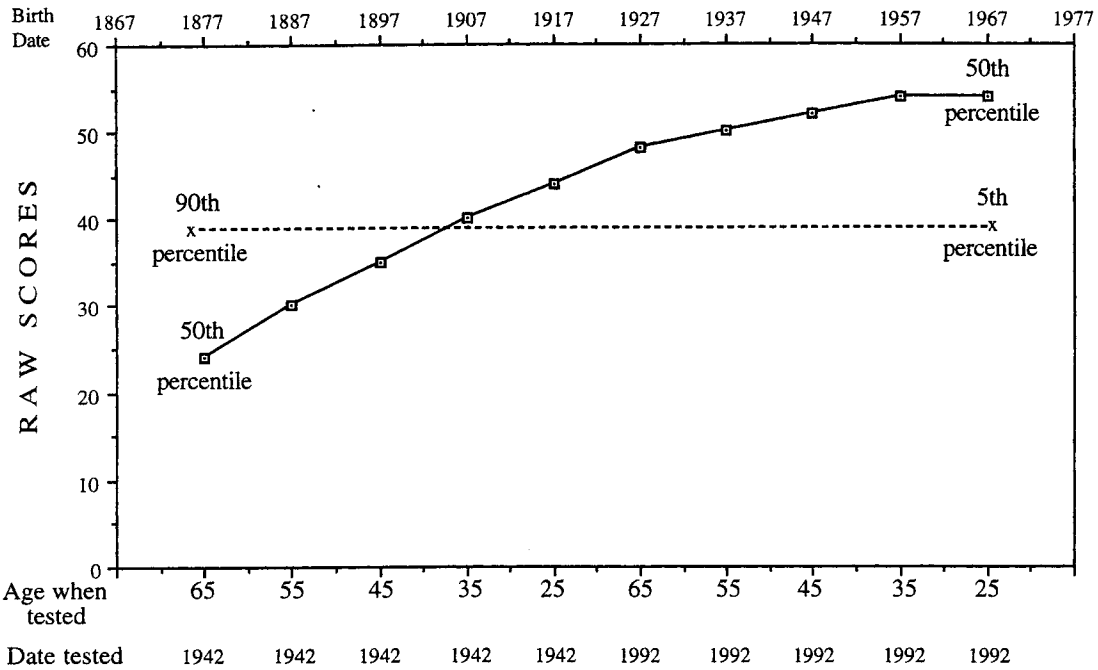
norms. The birth-date method overestimates gains because performance on Ravens declines in old age, and lesser sample quality dictates caution. Nonetheless, it is difficult to defend any estimate that puts less than 70% of late 19th century Britons below an IQ of 75 in terms of current norms. This deals a stunning blow to our confidence in the ability of IQ tests to compare groups for intelligence, at least when those groups are separated by cultural distance. Can anyone take seriously the notion that the generation born in 1937 was that much more intelligent than the generation born in 1907, to say nothing of the generation born in 1877? It also deals a blow to the Spearman–Jensen theory of intelligence. That theory is based on *g*, the general intelligence factor derived from the tendency of the same people to excel on a wide range of IQ tests. Moreover, it nominates the Ravens test as a marker test for *g*, that is, as a criterion by which the *g* loadings of other tests can be evaluated (Flynn, 1987c; Jensen, 1998, p. 38).

The contention that IQ gains cannot be identified with intelligence gains has been exhaustively discussed without consensus (Neisser, 1998). Rather than rehashing once again whether IQ gains should have been accompanied by concrete evidence of enhanced achievement in areas like academic performance, inventions patented, and so forth, I rest my case on what really convinces me: examples drawn from everyday life. Throughout my own life, the fact that height increases have been real has been only too evident. Going from being taller than almost all of my students to being shorter than so many has not escaped my notice. Have most of those who began life as clearly above average in intelligence experienced being overtaken by the majority of youth? Take a woman with an IQ of 110 who taught for 30 years in the Netherlands. As Figure 1 shows, during that period, the Dutch gained 20 IQ points on Ravens. So in 1952, she was brighter than 75% of her senior students; by 1967, they were her equals; by 1982, 75% of them were brighter than she was. Has that really been the career experience of Dutch teachers?

What conclusions must we draw about previous generations? Did most of them suffer from mental retardation? Jensen (1981, p. 65) emphasized the limitations of those with low IQ. He noted that someone with a Wechsler IQ of 75 or below may be a keen baseball fan and yet be vague about the rules, unsure of how many players make up a team, unable to name the teams his or her home team plays. How reasonable is it to assume that 70% of late 19th century Britons could not, even if it were their chief interest, understand the rules of cricket? The military data,

Figure 2

British Scores on Raven's Progressive Matrices: The Bottom 90% Born in 1877 Fall Within the Bottom 5% Born in 1967 (IQ 75 or Below)



Note. Adapted from Graph G2 of J. Raven, J. C. Raven, and J. H. Court, *Manual for Raven's Progressive Matrices and Vocabulary Scales*, Section 1, General Overview (1998 ed.). Oxford, England: Oxford Psychologists Press.

which are of impeccable quality, pose the same question. Can we assume that in 1952, almost 40% of Dutch men lacked the capacity to understand soccer, their most favored national sport? It may be objected that Jensen spoke of a Wechsler test rather than Ravens. However, we now have Wechsler and Stanford–Binet data that cover the period from 1918 to 1995 and show that White Americans gained almost 25 IQ points on this kind of test (Flynn, 1984, 1993, 1998h; Terman & Merrill, 1937, p. 50; Yerkes, 1921, pp. 654, 789). Does that mean that during World War I, almost half of White Americans could not understand the basic rules of baseball?

The international data fall into the same pattern as the American data. Gains are about 18 IQ points per generation (30 years) on Ravens, somewhere between 9 and 18 points on Wechsler and Stanford–Binet tests, about 9 points on purely verbal tests, small or nil on Wechsler subtests such as Arithmetic, Information, and Vocabulary (at least in English-speaking nations; German-speaking ones show substantial Vocabulary gains), and small or nil on achievement tests. That is to say, they fall away the closer we come to the content of school-taught subjects. This pattern is the opposite of what the Spearman–Jensen theory of intelligence predicts. Jensen (1988, 1989) has been experimenting with a battery of behavioral and physiological measures, reaction times or how quickly people can react to stimuli, average evoked potentials (AEPs) or the electrical response of the cerebral cortex to sights and sounds, plus the time an injection of glucose takes to reach and be absorbed by the brain. An ideal result would be that such a battery could replace IQ tests as a measure of intelligence where IQ tests give nonsense results. Reservations about this enterprise have been detailed elsewhere (Flynn, 1992, pp. 348–350, 1998c). My sole analysis of undigested reaction-time data showed that Chinese participants did better on measures that most correlated with Ravens for British participants, and that the British did better on measures that most correlated with Ravens for the Chinese. These results are discouraging, unless you believe that the Chinese have better British brains than the British do, and that the British have better Chinese brains than the Chinese do (Flynn, 1991b).

I believe that the failure of Ravens is an important piece of falsifying evidence. Some years ago, Jensen (1980) envisioned tests running from the detour problem through an adapted form of Ravens, which would allow us to measure the intelligence of cats and chickens, Kalihari Bushmen and Polar Eskimos, even extraterrestrials. Today we know that Ravens cannot bridge the gap between the Dutch of 1982 and the Dutch of 1967. There may have to be a radical rethinking of the Spearman–Jensen theory.

This is not to say that a theory that explains much is to be set aside lightly. Some theories hardly get off the ground. Descartes thought that as the sun spun on its axis, it created a whirlpool effect that propelled the planets into roughly circular orbits. Newton offered a far more fecund theory: It also had the planets moving around the sun but in accord with universal gravitation. The history of the Newtonian theory teaches an important lesson, namely, that the theoretical significance of a piece of falsifying evidence often emerges only after a considerable period of time. It predicted a stellar parallax: As the earth moved, the angle of observation from the earth to a fixed star ought to shift. No shift was found, which seemed shattering. It also predicted an orbit for Mercury that showed very minor discrepancies with the observed orbit. But this seemed a very minor problem, easily explained by observational error or the presence of a small inner planet. Yet, the absence of the stellar parallax was explained away (i.e., the stars turned out to be so inconceivably far away that their shift was unobservable with the instruments of the day), whereas Mercury's orbit eventually led to the radical reconceptualization that replaced Newton with Einstein. No one knows which role IQ gains over time will eventually play. Moreover, do recall that Newton has been transcended rather than discarded: His equations are a special case of Einstein's theory and are still useful for a wide range of practical purposes (Flynn, 1987c).

Jensen (1998, pp. 328–330) acknowledged that IQ gains pose theoretical problems, and he and I are not as far apart on this as he thinks. I have one reservation about his choice of words. It is not his recent decision to abandon the word *intelligence* in favor of *mental ability,* although because this achieves nothing, I will not follow suit (except in this paragraph). It is his use of *g* to refer to something beyond the mental ability that enhances performance on IQ tests. At times, Jensen speaks as if absence of gains in real-world cognitive abilities logically implies that there cannot have been *g* gains (Flynn, 1987c, p. 31; Jensen, 1980, p. 248, 1998, p. 329). I prefer to say that IQ gains are prima facie gains in terms of psychometric *g* and to discuss whether they have a significance too limited to signal an increase in real-world mental ability. IQ gains are prima facie *g* gains in the following sense: We have huge gains on tests (Raven and Wechsler) that Jensen has described as almost pure measures of *g*; indeed, factor analysis shows that they involve no significant factors beyond *g*. Let someone explain exactly how one generation can outscore another by 15 or 18 points on these tests and yet not be superior for psychometric *g* to virtually the same extent. I take up the prospects for such explanations in a moment. But assume they are not forthcoming and assume little evidence of the real-world achievements implied by identifying these psychometric *g* gains with mental ability gains. Then it must be crystal clear that such a failure counts against the theory: against any theory that asserts such an identity. To equate *g* with real-world mental ability obscures the fact that this equation is precisely what is under challenge. It is no more helpful than if Descartes had persisted in referring to the real-world forces that propel the planets through the heavens as "the whirlpool effect," even though it was a whirlpool theory that was being challenged by the evidence. All in all, it would be best to confine the term *g* to the theory's construct and never to use it to label some real-world cognitive ability.

The obvious path to defending the theory is to find the causes of massive IQ gains over time. Some causes, like test sophistication, or mere enhanced skill in taking tests, would neatly disentangle IQ gains from psychometric *g* gains and certainly would lead to no expectation of significant real-world effects in terms of intelligence. I believe that causes like test sophistication are nonstarters for reasons given elsewhere (Flynn, 1998c). Better education over time is often presented as a cause of IQ gains that would not enhance psychometric *g*. This would be true only if schools or parents were coaching children on how to take Ravens or were doing a better job of teaching relevant subject matter; for example, better teaching of arithmetic might enhance performance on the WISC Arithmetic subtest. There is no evidence for the former, and recall that it is precisely the subtests with school content like Arithmetic that show either nil or small gains. However, what if parents or schools are enhancing the generalized problem-solving ability that improves performance on IQ tests? Well that would be equivalent to enhancing *g*—not some poor substitute for *g*. In a word, it would simply mean that *g* can be taught, and is *g* that is taught by schools or parents somehow cursed in a way that *g* taught by less identifiable tutors is not? Taught *g* poses the same problem with the same force: Why hasn't enhanced *g* led to the full range of real-world achievement we expect from enhanced intelligence?

Patricia Greenfield (1998) and Ulric Neisser (1998) have suggested such things as video games and computers as possible causes. These factors loom large because gains on Raven's Progressive Matrices are so large, and matrices tests are sometimes suspected of having a significant spatial–visualization or perceptual organization factor. However, a comparative analysis yet to be published (Flynn, 1998e) shows that gains on the Similarities subtest of the WISC are the largest gains of all, running at perhaps 20 IQ points per generation. Like Ravens, Similarities is a test of fluid *g*. It tests for on-the-spot problem solving, rather than for the achievements characteristic of intelligent people, achievements like a large vocabulary and a large store of general information. Unlike Ravens, however, it has a purely verbal format and cannot be suspected of having a significant spatial factor. It is also worth noting that Ravens gains were large before the television era, much less before the computer-game era.

To briefly indicate the causal problem's complexity, a few words about the nutrition hypothesis. Richard Lynn (1987, 1989) believes that at least 15 points of IQ gain over the last 50 years can be explained thereby, and he points to height gains of similar magnitude. I present three points in rebuttal. First, some European countries have gained two or three standard deviations of height over the last century or two (Floud, Wachter, & Gregory, 1990, pp. 16, 23, 26). If height gains are truly accompanied by intelligence gains,

they pose a familiar question: Did the Dutch in 1864 really have the same intelligence as people who today score 65 on IQ tests; did Norwegians in 1761 really resemble those who today score 62? Second, experimental study of the effects of vitamin–mineral supplements on IQ has shown in California that a modest supplement had little effect, a moderate one had significant effect, and a large one little effect (Schoenthaler, Amos, Eysenck, Peritz, & Yudkin, 1991, pp. 357–358). That every nation has continuously enhanced nutrition just the right amount, neither too little nor too much, for decade after decade, seems unlikely. The Netherlands almost certainly gave children born after the Second World War better nutrition than it gave those born during the great wartime famine. The effect of this change in nutrition on IQ gains over time was nonexistent (Flynn, 1992, p. 346). Third, if nutrition has caused IQ gains of 15 or 30 or 45 points, we seem driven to posit huge understanding-baseball intelligence gains. After all, better nourished brains should function better in everyday life, not just in the test room, yet intelligence gains of this magnitude seem absurd. In other words, anyone who wants to solve the causal problem must explain *why* massive IQ gains have occurred—*without* implying that those gains should be attended by intelligence gains.

In my opinion, counterexamples cast doubt on the causal hypotheses advanced thus far: increased outbreeding, test sophistication and its variant the Brand hypothesis (Brand, 1987a, 1987b), the nutrition hypothesis and its variant the Storfer hypothesis (Storfer, 1990), enhanced socioeconomic status (SES), urbanization, more or better education, and the advent of television (Flynn, 1998c). Then there is the global hypothesis that the sheer complexity of the modern world is responsible (Vincent, 1993, p. 62), but global hypotheses cannot be tested unless mediating variables are spelled out. Schooler (1998) has contended that occupations and housework have become more cognitively demanding. On the other hand, Sowell (1972) concluded that the overwhelming bulk of jobs are not more intellectualy demanding and that many have become simplified to the point of boredom. Let someone show that driving and maintaining a car, or running a farm, or managing a home, or interacting with kin foster cognitive skills today that were absent yesterday and that these match the skills demanded by Ravens.

Wendy Williams (1998) has done a good job of listing virtually every trend that might qualify as a plausible cause given the total pattern of IQ gains. Some of these look promising, but they must be assessed in terms of the full complexity of the problem at hand. Hypotheses must perform at least three tasks. They must explain the duration, magnitude, and pattern of IQ gains, account for their very limited or at least highly compartmentalized real-world effects, and explain why environmental factors are so potent between generations and yet so feeble within generations. I (Flynn, 1998e) suggest a change in strategy: Why not use score patterns to simulate three generations? Doctoral candidates at 10 universities (volunteers welcome) could select a large sample of children and divide them into three groups. The first would be average for WISC Arithmetic, Information, and Vocabulary but inferior for Block Design, Similarities, and Ravens, the score pattern of the previous generation. The second would be average on all, the score pattern, of course, of the present generation. The third would be average on Arithmetic, Information, and Vocabulary but superior for Block Design, Similarities, and Ravens, the score pattern of the next generation (we think). Then we could study how they differ in attitudes, character, hobbies, schooling, family structure, number of siblings, whatever we want. For the first time, we might be blessed with causal hypotheses with a solid evidential base.

Achievement Beyond IQ

Why go from researching IQ gains over time to researching the IQs and achievements of Asian Americans? Jensen provided the bridge. He did so by using an analogy to illustrate the significance of IQ gains over time (Jensen, 1992, p. 277). He imagined people using shadows to measure height. If they use them to make comparisons over time—if they compare shadows at high noon with shadows late in the day—the latter are longer and register height gains that are spurious. The distorting factor to which shadows are sensitive is, of course, the change in the angle of the sun's rays. However, at a particular time and place, that distorting factor will be absent, and shadows will rank people for height with considerable accuracy. Jensen's analogy has an interesting implication. Let us assume that IQ is defective in comparing intelligence for groups that are separated by time. Nonetheless, IQ might tell us something very significant about groups living in the same place at the same time. Take America's ethnic minorities. The present generation of an ethnic minority does not compete with a previous generation of American Whites. It must compete with its White contemporaries in the context of a particular social structure. Jensen (1973b), Gottfredson (1987), and Nichols (1987, p. 215) have all emphasized the potency of the mean IQ of ethnic groups in determining their real-world achievements. Nichols believes that IQ determines a group's relative position in terms of occupation, income, family demoralization, and crime. How much of this is true? Does their mean IQ really chain ethnic groups to their particular pair of oars in American society?

In *Asian Americans: Achievement Beyond IQ* (Flynn, 1991a), I analyzed the published literature on Chinese American and Japanese American IQs between 1938 and 1985. The focus is on the generation of individuals who were born in the late 1940s, graduated from high school in the mid-1960s, and attained their adult occupations by the 1980 census. Ten studies, covering about 7,500 participants, were deemed reliable and relevant, and these confirmed the values for 12th graders from the 1966 Coleman Report (Coleman, 1966). Normed against a mean of 100 and a standard deviation of 15 for Whites, Chinese and Japanese Americans received 100 for nonverbal IQ, 97 for verbal IQ, and 98.5 for overall IQ (Flynn, 1991a, Tables 4.1 and 4.2). In other words, they did no better than match Whites. This assertion applies only to the generation in question. That generation was raised in homes whose SES matched that of White Americans. Their children, thanks to

the achievements of their parents, were raised in high-SES homes, and after 1949, the ranks of Chinese and Japanese Americans were swelled by an elite immigration. The current generation probably has a mean IQ above that of White Americans.

As for the achievements of the postwar generation, Chinese Americans opened up large "IQ–achievement gaps." An IQ–achievement gap is a measure of the extent to which a group has transcended the limitations of its mean IQ. Recall that these Chinese Americans had a mean IQ of about 99 when normed against Whites. However, they were so high on the U.S. occupational hierarchy that they matched a subgroup of the White population with an IQ of 120. That is, they performed as if they had a mean IQ 21 points higher than they actually had, which means their occupational IQ–achievement gap measured at 21 points. Japanese Americans enjoyed a gap of 10 points. For income, the two groups were closer together, with gaps of 16 and 14 points, respectively (Flynn, 1991a, Tables 4.4 and 4.5).

A method was designed to partition IQ–achievement gaps. The 21-point occupational gap for Chinese Americans was divided into two components. About 7 points were from a *threshold factor*: This means that Chinese Americans could spot White Americans 7 IQ points and still get the academic credentials needed to qualify for elite occupations. About 14 points were from a *capitalization factor:* This means that Chinese Americans capitalized on 78% of their pool of potential qualifiers, whereas White Americans capitalized on only 60% of theirs (Flynn, 1991a, Tables 5.1 and 5.8). In other words, Chinese Americans with IQs of 93 and above could qualify for elite occupations, and 78% of those actually did. Whites needed an IQ of 100 or above to qualify, and only 60% of those actually did. The percentage that realize their potential has much to do with characterological traits. Take an Irish American and a Chinese American, both of whom have qualified for Stanford Law School, and both of whom have a fiancée who wants them to stay at home. The Irish American is more likely to stay at home, whereas the Chinese American is more likely to get a new fiancée.

The high fliers, both within U.S. society and without, are those groups that soar beyond the assumed limitations of mean IQ. According to data from Stevenson et al. (1985), Japanese children in Japan open up a huge IQ–achievement gap when compared with White children in America, fully 19 points between nonverbal IQ and school-taught arithmetic. Chinese children in Taiwan open up a 16-point gap. According to Project Talent data (Backman, 1972), Jewish Americans turn a 10-point deficit for visuospatial IQ into a 10-point advantage for high school mathematics. Their overrepresentation as mathematicians and statisticians yields a 30-point gap.

The achievements of Jewish Americans falsify the frequently heard gender hypothesis that women must lag behind men in mathematics because they suffer from a visuospatial deficit. Jewish Americans get high scores for verbal IQ. Perhaps that compensates for their visuospatial deficit by allowing them to think their way through mathematical problems. In fact, no mix of verbal and visuospatial IQ measures can eliminate huge IQ–achievement gaps for various combinations of ethnicity and gender (Flynn, 1991a, pp. 116–125). Lynn (1994) has suggested a gender hypothesis that would mean a total reevaluation of women's achievements: the hypothesis that women suffer from an overall IQ deficit of about four points. I (Flynn, 1998d) analyzed Israeli military data that included virtually all men and 80% to 85% of women. The data for women showed sufficient variation to yield excellent projections for the total population. Israeli women were about 0.8 points above men on a Verbal IQ test and perhaps 1.4 points below men on a matrices test: When these are averaged, they give an overall difference of only 0.3 points.

Calculating the mean IQ of Chinese and Japanese Americans illustrates the practical implications of IQ gains over time. As we have seen, the standardization samples used to norm IQ tests keep improving their performance from one decade to another. Therefore, individuals or groups merely average in their own time will appear above average if scored against a test normed 10 or 20 or 30 years before. IQ gains over time create obsolete norms, and scoring against obsolete norms gives inflated IQs. In his great book on Asian Americans in North America, Vernon (1982) based his summary estimate of Chinese American IQ on samples scored against the obsolete norms of a Lorge–Thorndike test standardized 22 years earlier; therefore, he gave an inflated estimate, at least for that generation of Chinese Americans, and the fact that they had achieved far beyond their mean IQ was overlooked (Flynn, 1991a, pp. 49–51; Vernon, 1982, pp. 23, 28). The effects of the massive intervention called the Milwaukee Project could only be evaluated after the children's IQ scores were corrected for obsolescence. These corrections showed the children with a mean IQ of 92 at age 14, a mean about 9 points higher than that of the control group. Achievement test scores confirmed that 9 points was the true cognitive gain (Flynn, 1998g, pp. 36–37). The famous Skodak and Skeels adoption study found a 20-point IQ gap between adopted children and their biological mothers, but this reduced to some 10 to 13 points after correction for obsolescence. Adoptive homes were getting credit for illusory cognitive gains (Flynn, 1993).

The day after an IQ test is published, IQ gains begin to diminish the number of people who score below a score of 70. Those who rise above that score are no longer eligible to be classified as suffering from mental retardation. Then when the test is renormed, all of those who have escaped are of course pulled back into the pool. The past 50 years have seen the standardizations of the WISC, the WISC-R, and the third edition of the WISC (WISC-III; Wechsler, 1992). Although a change in the criterion of mental retardation was a contributing factor, the combination of IQ gains and the obsolete norms they engendered were the primary cause of huge fluctuations in the proportion of children eligible to be so classified. Indeed, the proportion has ranged from a high of 1 in 23 (1949) to a low of 1 in 213 (1989). These fluctuations brought no response from practicing psychologists: They seem to have

been unaware that whenever a test was renormed, their IQ criterion began to label thousands of children as mentally retarded who had escaped the label the previous day. So much for the notion that we have ever had a consistent IQ criterion of mental retardation or that any criterion had a coherent consensus of clinicians in its favor (Flynn, 1985; in press-a).

Despite current awareness of the problem of obsolete norms, mistakes continue to be made. Sometimes diagnoses are made by comparing a child's performance on two or more tests. For example, the recent WISC-III (Wechsler, 1992, pp. 212–213) manual gives a criterion for learning disabilities or reading disorders in terms of differential performance on four WISC subtests. This assumes that IQ gains will do no harm because, although the norms may become obsolete, children will still be scored against subtests that were normed at the same time—and therefore, against subtests equally obsolete. In fact, WISC subtests show radically different rates of gain. Therefore, some 15 years after a WISC test is published, the scores of perfectly normal children may be inflated by different amounts on the relevant subtests. Therefore, they would be in danger of misdiagnosis (Flynn, in press-a).

IQ Gains and Race

IQ gains over time provide a wonderful laboratory to test many of the hypotheses put forward in the race and IQ debate. The best way to illustrate their potential is to divide the history of that debate into the pre-IQ-gains era and the post-IQ-gains era. I was a late entry in this debate because only in 1977 did I become aware of Arthur Jensen's seminal 1969 article in the *Harvard Educational Review* (Jensen, 1969). Jensen argued that even if environments were equalized, the 15-point IQ gap between American Whites and American Blacks would only be reduced to something like 10 points. Having been a chairperson for CORE (Congress of Racial Equality) in a Southern state at the time of Martin Luther King (Flynn, 1967, chap. 6–7), I believed that many Blacks suffered from low self-esteem. I was certain that they were discriminated against because of their group membership. Therefore, Jensen's thesis was unwelcome. It was a blow to Black pride. It implied long-term below-average group achievement by American Blacks. This is significant because even unbiased people are likely to judge individuals by group performance rather than by personal traits (Flynn, 1996). However, no one has the right to attack scholars simply because their research has led them to unwelcome conclusions. The truth cannot be racist, nor can anyone be held suspect for telling the truth as they see it, unless their assessment of evidence falls below the minimum level we expect of a competent scholar. Jensen's case was clearly too powerful to fall into that category.

A short digression is relevant. If Jensen's thesis is correct, the path to social justice will be more difficult. However, if he is correct, he has done us the favor of forcing us to face a facet of reality many are reluctant to accept. Would anyone who holds humane ideals prefer to pursue them in a fantasy world rather than the real world? The reluctance to give Jensen a fair hearing is proven by the popularity of bad arguments designed to make the genetic equality of groups an undiscussable dogma. The usual ones are that the concept of race is not biologically respectable, that all human groups share most of their genes, that the concept of intelligence is culturally relative, and that current theories of intelligence have not given an adequate pretheory definition of the term. Brief rebuttals to these arguments run as follows: The races investigated are defined sociologically. Despite their similarity, human groups show enough genetic variation to cause statistical differences for other traits, like occurrence of sickle-cell anaemia, so why not intelligence? Black parents want their children to excel in the kind of intelligence that pays dividends in America or England or France, not in some preindustrial society. Finally, even the hardest sciences do not give elaborate pretheory definitions of their key concepts. Newton did not wait to refine the concept of celestial influence before embedding it in his theory of gravitation (Flynn, 1987c, 1994a).

Race, IQ and Jensen (Flynn, 1980) was an attempt to confront Jensen's thesis about Black and White on purely logical and evidential grounds. Jensen had put forward a two-step case. First, he argued that only 12% of IQ variance within White Americans was explained by between-family environmental factors, and he suggested that probably much the same was true for Black Americans. This was disturbing because the environmental factors that separate families are things like income, housing, schooling, professional versus working class home life—the kinds of things that are captured when we speak of differences in SES—and we were being told that their effect on IQ differences was relatively slight. Second, he argued that the factors environmentalists nominate to explain the Black–White IQ gap simply could not account for the full 15 points. For example, when someone first hears that Black IQ is 15 points below White IQ, the typical reaction is to ask what happens if we match the races for SES. Rather than eliminating all or most of the IQ gap, matching for what seems the most potent and all-embracing factor eliminates only 3 points (Jensen, 1972, 1973a, 1973b).

Jensen (1972, 1973a, 1973b) then made a critical observation. He noted that the factors environmentalists suggest to explain the IQ gap *between* the races are very similar to those they use to explain differences *within* each race. SES is a perfect example. This observation is critical because it allows Jensen to treat Blacks as if they were a subgroup or sample of the White population. After all, if the environmental factors that divide Blacks from Whites are the same as those that divide Whites from Whites, then Blacks are like poor Whites, similar to a group that happens to be environmentally underprivileged within the White community. This allows Jensen to combine his first contention and his second contention and to achieve a logical force beyond what either would provide in isolation. Now an environmental hypothesis about the racial IQ gap can be tested as follows: If purely environmental factors selected out a subgroup of the White population, how much envi-

ronmental deficiency would be necessary to account for a 15-point (one standard deviation) IQ deficit?

The key to what follows is the concept of correlations as measures of regression toward the mean. To provide a simple example: Imagine that the correlation between height and between-family environment was perfect, or 1.00. The significance of that would be this: If we found a group one standard deviation below the mean for height, and environmental factors were solely responsible for their height deficit, then they should be one standard deviation below average in terms of environment. Now imagine the correlation is less than perfect, perhaps only .35. In that case, it would take several standard deviations of environmental deprivation to account for a one standard deviation height deficit. A bit of arithmetic shows it would take 2.86 standard deviations, because 2.86 times .35 equals one standard deviation. In summary, the correlation determines precisely how many standard deviations of environmental deficit it takes to account for a one standard deviation height deficit, or to be technical, it determines how far toward the mean the below-average group will regress as each standard deviation of environmental deficit is eliminated: clearly only 35% of the way.

What Jensen needs to complete his argument is the correlation between between-family environment and IQ. Recall the estimate that 12% of IQ variance is explained by between-family environment. Thanks to the mathematics of a normal curve, the square root of that value will give him the desired correlation: The square root of .12 equals .35; therefore, that is the correlation between between-family environment and IQ. Following the above example about height, we now know precisely how many standard deviations of environmental deficit it would take to account for a one standard deviation IQ deficit: The average environment of Blacks would have to be 2.86 standard deviations below the quality of the average environment of Whites. That is a huge environmental deficit. The average environment of Black Americans would have to be inferior to that enjoyed by 99.79% of White Americans. Can anyone maintain that such a thing is plausible?

Finally, Jensen finishes us off by offering us a poisoned apple, an escape that looks attractive but proves fatal. He says we can always posit a mysterious factor *X*. This factor *X* would have to be an environmental factor, hitherto unidentified, that is potent between the races and virtually uniform within the Black population. It must be so potent that it costs every Black individual about 15 IQ points but so uniform that it affects no Black individual much more than any other, otherwise, it would have a potent differential effect on the Black population and would raise the proportion of between-family environmental variance for Blacks way above 12% (Jensen, 1972, 1973a, 1973b).

The apple appears attractive because there is an obvious candidate for factor *X*. Racism looks like a potent environmental factor that affects all Blacks both negatively and with considerable uniformity. However, as said previously, global explanations are suspect without the identification of mediating variables that link them to the phenomenon they are supposed to explain. Racism is not some magic force that harms Blacks without a chain of causality. Racism harms Blacks because of certain effects, such as lack of confidence, low self-image, emasculation of men, the welfare-mother home, poverty. Who could argue that these same factors do not vary significantly within the Black population? Certainly, Black Americans are divided by high and low self-esteem, stable and unstable homes, high and low incomes, good and bad housing. If these factors both are potent and vary among Blacks, why do they explain so little IQ variance within the Black population?

Lewontin (1976a, 1976b) challenged Jensen's analysis by presenting a counterexample, that is, a case in which genes would explain virtually all variance within each of two populations, but environment would explain an average difference between them. He imagined a sack of seed corn with plenty of genetic variation. The corn is randomly divided into two batches, each of which will therefore be equal for overall genetic quality. Batch A is grown in a uniform and optimal environment, so within that group all height differences at maturity are due to genetic variation; Batch B is grown in a uniform environment that lacks enough nitrates, so within that group all height differences are also genetic. However, the difference in average height between the two groups will, of course, be due entirely to the unequal quality of their two environments. At the time, Lewontin's counterexample struck me as simply an evasion. Jensen never denied we could construct a theoretical model of groups equal in genetic quality, one of which was afflicted by a factor *X* that handicapped all of its members uniformly and affected no one else. His challenge was to suggest some real-world example that matched such a theoretical model, without clearly exceeding the bounds of plausibility.

Unable to provide, or even to imagine, such a real-world example, my own critique tried to render his analysis impotent by going to the core of the issue. It rested on the distinction between direct and indirect evidence. Direct evidence comes from situations where Blacks actually exchange their usual environment for one that approximates a White environment or at least a neutral environment. Indirect evidence is evidence that, in the present state of our knowledge, looks relevant to predicting what would happen if Blacks were to make that kind of environmental exchange. Matching Blacks and Whites for SES is an example of indirect evidence. It is an attempt to predict what would happen to IQ among Blacks if their economic disadvantages were eliminated. Jensen's analysis is another example: It predicts that no reasonable degree of environmental progress could bring Blacks to IQ parity with Whites. However, in terms of logic, indirect evidence must give way to direct. Assume that whenever Blacks are reared by Whites in an environment normal for Whites, they achieve IQ parity; then there would have to be something wrong with Jensen's analysis.

In *Race, IQ, and Jensen* (Flynn, 1980), I argued that the direct evidence favored an environmental hypothesis. It conceded one exception: The Minnesota Adoption Project, in which White parents had adopted at least one non-White

child, was analyzed as a piece of direct evidence perhaps unfriendly to an environmental hypothesis. This was because adopted children who had two Black parents scored 12 IQ points below children who had one Black and one White natural parent (Flynn, 1980, p. 104). Scarr, Weinberg, and Waldman (1993) have updated the debate about the significance of transracial adoptions.

However, there was and is a powerful piece of direct evidence in favor of genetic equality. The soldiers of the American occupation force in Germany, both White and Black, fathered thousands of children with German women after World War II. Eyferth (see Flynn, 1980) selected a representative sample of 181 Black children and a matching group of 83 White children and found that their mean IQs were virtually identical. In other words, children who had a White father seemed to possess no advantage whatsoever over those who had a Black father. This is significant only if these Black and White soldiers were representative of the larger American populations of Black and White men. Analysis of the total body of army mental test data, held at Suitland, Maryland, showed that the White soldiers were an elite by one point of heritable IQ; the Blacks were an elite by about two to four points. Therefore, 80% to 90% of the racial IQ gap should have been present in the genes of the fathers. The fact that it was absent in their children would indicate that the gap is almost entirely caused by environment. Even the small remainder may be explained away in that Eyferth believed the Black children suffered a special handicap because their color advertised their illegitimacy (Flynn, 1980, pp. 84–102).

It was also necessary to test hypotheses about selective mating, for example, that intelligent Blacks possessed an advantage in securing sexual partners. The army thought it worth spending time and money to investigate virtually everything in terms of race. Their records suggest that it was the propensity to spend freely that conferred an advantage, and that Blacks who were sexually active actually scored below the Black mean on the Army General Classification Test (Flynn, 1980, pp. 94–95). Eysenck (1981) hypothesized that the Black children may have had a certain advantage, namely, that racial admixture might confer the benefits of hybrid vigor. At the other extreme, some suggest that racially mixed offspring may suffer from some sort of reproductive stress, and that this might have an adverse effect on IQ (Loehlin, Lindzey, & Spuhler, 1975). In my view, American Whites and Blacks are both already so hybridized that evidence from animal hybrids has no true counterpart.

In England, Black children admitted in infancy to long-stay residential nurseries outscored White children therein (Tizard, 1974; Tizard, Cooperman, Joseph, & Tizard, 1972). Allowing for selective migration from the West Indies to England drops Blacks down to parity with Whites (Flynn, 1980, pp. 108–113). Most of these children were only four at age of testing, and no follow-up has been done. It would be of limited significance, of course, because the participants have lived almost all their lives within the Black environment of Britain. The other studies that might qualify as direct evidence, ranging from the studies of Witty and Jenkins (1934, 1936) to the blood-group studies (Loehlin, Vandenberg, & Osborne, 1973; Scarr, Pakstis, Katy, & Barber, 1977), seem to show that Blacks derive no benefit from a higher-than-average degree of White ancestry, but all contain serious methodological flaws (Flynn, 1980, pp. 75–84, 262–264). Mackenzie (1984) has described a research design that might yield valuable results if used.

Direct evidence takes priority, but there is not much of it. Therefore, it would be a mistake to leave the indirect evidence unassessed, particularly Jensen's powerful two-step case. Recall its core: High heritability of IQ within groups implies that genes are strong and environmental factors weak therein. The case for an environmental explanation of large between-group IQ differences has degenerated into positing a mysterious factor *X*, a factor potent between groups but feeble within. No one has been able to find a real-world example that shows the footprints of a factor *X*. Failing that, the environmental gap one must posit to explain an IQ difference of even 15 points (one standard deviation) is impossibly large.

The pre-IQ-gains literature offered little by way of rebuttal. The best that could be done was to try to show that within-group heritability was somewhat lower than Jensen believed and to suggest a few candidates for factor *X* that might at least aspire to the role (Flynn, 1980, chap. 4–5). This was at best half-convincing. However, the discovery of IQ gains over time ushers in a new era. IQ gains can do three things: provide real-world examples that show the footprints of a factor *X*, reduce the environmental gap one must posit to explain the Black–White IQ gap to a reasonable size, and allow us to test supplementary evidence that has been aduced in favor of a genetic hypothesis in recent years. These three claims are defended ad seriatim.

Several modern industrial nations show IQ gains on Raven's Progressive Matrices averaging at 18 points over 30 years; several show gains on Wechsler Performance Scales almost as great; the latter also show gains on the Similarities subtest from the Verbal scale averaging at 20 points. These nations give us two groups, the present generation and the previous generation, with a between-group IQ gap larger than that between Black and White Americans. The literature that partions the IQ variance within each generation is not comprehensive, but it suggests that the American values that favor genes over environment also hold for most industrialized nations (Bouchard & McGue, 1981; Jensen, 1998, p. 198; Loehlin, 1989; Plomin & Rende, 1991). Despite this, the IQ gap between the generations must be entirely environmental. Indeed, if those who tell us dysgenic trends are at work are correct (Lynn, 1996), environmental factors have to swamp negative genetic factors and have a potency greater than 18 or 20 points.

So now we have a real-world example that fits Lewontin's (1976a, 1976b) theoretical model: huge between-group IQ differences that are environmental in origin despite genes dominating the explanation of within-group differences. Therefore, either the apparently absurd factor *X*s are actually quite common, operating here and there and

everywhere, or something is badly wrong with the methods used to partition IQ variance between genes and environment. The link between within-group and between-group genetic factors has been broken. Perhaps this was to be expected: We have always known that within-group heritabilities were useless in predicting which group might have the higher mean. Therefore, how relevant could they be? Apologies are due to Lewontin and were rendered a decade ago (Flynn, 1989, pp. 365–366).

As for the environmental gap one must posit to explain the Black–White IQ gap, IQ gains over time pull this out of the stratosphere and down to earth. It appears that Blacks have enjoyed a slightly higher rate of gain on Wechsler-type tests than Whites (Herrnstein & Murray, 1994, pp. 277, 289). This implies that since 1945, Blacks have gained at an average rate of over 0.30 points per year and have gained a total of 16 points over 50 years (Flynn, in press-b). Therefore, the Blacks of 1995 should have matched the mean IQ of the Whites of 1945. Therefore, an environmental explanation of the racial IQ gap need only posit this: that the average environment for Blacks in 1995 matches the quality of the average environment for Whites in 1945. I do not find that implausible.

This brings us to new arguments that supplement Jensen's original two-step case. The most powerful focus on certain correlations. If you rank the 10 subtests of the WISC for inbreeding depression, and then rank them for the magnitude of their Black–White IQ gaps, you get a positive correlation of about .26. If you rank the WISC subtests for their *g* loading, the positive correlation with the Black–White IQ gaps may be higher at about .43 (Rushton, 1998, Table 2). Inbreeding depression measures the extent to which IQ is depressed by interbreeding among close relatives, which is known to have deleterious genetic consequences. The *g* loading of tests refers to the general intelligence factor, which is known to have a robust genetic component within groups. It seems disturbing that the worse Blacks do on WISC subtests, the more those subtests load on things like inbreeding depression and *g*. Jensen (1997) is guarded about the significance of these correlations. However, Rushton (1997) believes that they constitute a method that can diagnose whether the Black–White IQ gap has a potent genetic component: Inbreeding depression is posited as a primary indicator for such a component; *g* loadings are posited as a secondary indicator. He takes Stephen J. Gould to task for ignoring this "critically important finding" in favor of a genetic hypothesis (Rushton, 1997, p. 176).

IQ gains over time can be used to test Rushton's method. Five data sets from four nations, all of the available data, were merged to rank the 10 WISC subtests for the magnitude of IQ gains over time, all the way from losses on Arithmetic to huge gains on Similarities. The Spearman rank-order correlation with the subtests ranked for inbreeding depression, when the subtest reliabilities were partialled out, was positive at .26. This matches the correlation Rushton found between Black–White IQ differences and inbreeding depression (Flynn, 1998f). So now we know that inbreeding depression is bankrupt as a primary indicator of whether group IQ differences are mainly genetic. It tells us that the IQ gap between the generations is genetic when it is known to be overwhelmingly environmental. It simply gives nonsense results (Flynn, 1998a).

Rushton (1998) then called attention to his secondary indicator of genetic causality: the positive correlation between *g* loadings and Black–White IQ gaps. He pointed out that there was a negative correlation between *g* loadings and IQ gains. Thus, the secondary indicator gave a sensible result: It at least diagnosed that IQ gains were environmentally caused. Moreover, he did a factor analysis that produced a cluster inclusive only of IQ gains on the one hand and a cluster inclusive of Black–White gaps plus inbreeding depression plus *g* loadings on the other. He argued that this salvaged the method: Perhaps it gave a nonsense result when it used inbreeding depression alone, but when both indicators of genetic causality were taken into account, IQ gains became separated from them and were revealed to be environmentally caused. That the factor analysis has that significance need not be conceded. It seems odd that the primary indicator gives a nonsense result and needs the secondary indicator to bail it out. However, a more direct rebuttal is available. The factor analysis gives a sensible result only if the correlation between *g* loadings and IQ gains really is negative. The *g* loadings used are those derived from the WISC battery itself, but what if there is an alternative *g* that gives a different result? Perhaps if we were to rank WISC subtests in terms of fluid *g*, rather than crystallized *g*, the correlation with IQ gains would change from negative to positive (Flynn, 1998f).

This hypothesis gains plausibility from two things. The *g* calculated from the WISC battery has a crystallized bias (Jensen, 1987b, p. 96), and the pattern of IQ gains shows that they are primarily gains in fluid *g*. IQ gains are highest on Ravens and on the Similarities subtest of the WISC, and they are lowest on WISC subtests like Vocabulary and Information. Recall how these two kinds of tests differ. When asked "what do water and salt have in common?" (Similarities), you have to do on the spot thinking, actively imagine possibilities (both in ocean, both at meals, both edible), and choose the best one. That shows fluid intelligence. When asked "what is the capitol of Greece?" (Information), you either know that it is Athens, or you do not, and the response is automatic. That shows crystallized intelligence, demonstrates that someone has the vocabulary and information we would expect a bright person to accumulate by a certain age. In terms of current IQ-based theories of intelligence, fluid *g* plus a bit of motivation equals crystallized *g*, and crystallized *g* plus more motivation equals achievement. If you are intelligent, you do not need much motivation to become fluent, but you need considerably more to become a lawyer. As this implies, fluid *g* is a more factor-pure measure of intelligence than crystallized *g* (Jensen, 1980, p. 647). Therefore, no apology is needed for using fluid *g* to rank WISC subtests.

How can that be done? Jensen (1998, p. 38) provided a clue. He asserted that when the *g* loadings of tests within a battery are unknown, the correlation of Ravens with each test is often used to get an estimate. Because Ravens is the

recommended test of fluid *g*, it seemed worth determining the correlation of Ravens with each of the WISC subtests. Expectations were not high because among normal people, fluid *g* and crystallized *g* tend to go together (Jensen, 1980, pp. 234–235). It really is true that those with mental agility usually also develop large vocabularies and stores of general information. There are only five studies in the U.S. literature that give correlations between Ravens and the various WISC subtests. When pooled, these did affect the subtest hierarchy, but it was clear that Ravens had only half disentangled the two *g*s. For example, the loading of Ravens on the Similarities subtest was less than the loading of Ravens on Vocabulary. Given this, the results were astonishing. The Spearman rank-order correlation between the Ravens loading of WISC subtests and the magnitude of their IQ gains was about .50; when subtest reliabilities were partialled out, it rose to .60. Moreover, as numbers increased to the total of the 439 existent in the five studies, the correlation tended to rise. Preliminary analysis suggests that optimal numbers would boost the correlation to .70 (Flynn, 1998e). The correlation between IQ gains and fluid *g*, fully disentangled, must be very high indeed.

This completes the refutation of these supplementary arguments for a genetic IQ gap between Black and White Americans. Both of the indicators of genetic causality posited by the method give nonsense results. Both the correlation between IQ gains and inbreeding depression and the correlation between IQ gains and fluid *g* are positive. The method declares IQ gains to be mainly genetic; they are known to be mainly environmental.

The weakness of the indirect arguments examined underlines the message of not giving such arguments too much authority. When my critique of his two-step case was first published, Arthur Jensen was bemused by the gross exaggeration of the claim made in my title: "Jensen's Case Refuted" (Flynn, 1987d; Jensen, 1987a, p. 379). Indeed, it would have been absurd to think that simply because one argument for the thesis that there is a Black–White genetic IQ gap was suspect, the thesis had been disproved. The claim was more modest: I claimed only that Jensen's two-step case had been refuted; I questioned only that one argument, albeit a very important argument. Now, in my opinion, some more indirect arguments have been refuted. However, plenty of others have been put forward, and this brief review does not come close to criticizing them all. Nonetheless, given the track record of indirect arguments, given how often indirect evidence has seemed overwhelming, only to prove flawed after reflection, must we not wait for a solid body of direct evidence? The troops in postwar Germany provide the best example of Blacks actually escaping the usual American environment, and they provide evidence in favor of an environmental hypothesis. No one could consider it decisive, but it signals what we would need to reach a firm conclusion. The appropriate rejection of Black genetic inferiority is this: Nothing at present coerces rational belief.

Merit and Class

In 1994, Herrnstein and Murray published *The Bell Curve: Intelligence and Class in American Life.* They argued that affirmative action should be curtailed, partially on the grounds that racial bias has become negligible in America. The evidence for this appears anecdotal, but more important, the case for affirmative action need not rest either on the presence of racial bias or on the righting of historical wrongs. Elsewhere I have argued that completely unbiased rational actors in America will use color as a cheap information-bearing trait, and that Blacks will be judged by their group membership rather than as individuals. This profoundly disadvantages Blacks in terms of police behavior, access to jobs and housing and finance, and consumer prices. Perhaps most important of all, it prevents Black women from using marriage as an escape from the solo-parent poverty trap. The consequences amount to a systemic affirmative action program for Whites, which should be accompanied by a public affirmative action program for Blacks as compensation (Flynn, 1996).

However, this is a distraction from the real challenge *The Bell Curve* poses. The humane-egalitarian concept of social justice includes more than compensating people who suffer because of their group membership. It gives high priority to certain ideals, such as reducing environmental inequality and social privilege to tolerable levels. Herrnstein and Murray (1994) went beyond race to level the most devastating possible critique of those ideals, namely, that they self-destruct in practice. I refer to the meritocracy thesis, which runs as follows. The closer we come to environmental equality, the more all talent differences become caused by genetic differences. The more we eliminate privilege, the more we have total social mobility, and good genes for talent rise to the top and bad genes sink to the bottom. The tendency to marry those of similar IQ produces mating couples whose social status correlates with genetic quality. The result is an elite class whose children replicate their parents' high status, because of luck in the genetic lottery, and a large immiserated underclass whose children, handicapped by their bad genes, cannot escape low status. How little this vision will appeal will vary from person to person, but it is safe to say that countless idealistic men and women did not lay down their lives for this.[1]

Herrnstein and Murray (1994) used IQ as a marker for talent. They hypothesized that a potent meritocratic trend began in 1960, and this hypothesis generates a prediction. If they are correct, the IQ gap between upper and lower class children should show a visible jump when representative samples of children tested recently are compared with those tested in the pre-meritocratic era. The best data come from the White members of Standard–Binet and Wechsler standardization samples selected from 1932 to

[1] The refutation of the meritocracy thesis that follows is adapted from "Group Differences: Is the Good Society Impossible?" by J. R. Flynn, 1996, *Journal of Biosocial Science, 28,* pp. 580–584. Copyright 1996 by Cambridge University Press. Adapted with permission.

Table 1
Mean IQs of Children Whose Parents Were in the Top, Middle, and Bottom Thirds of the U.S. Occupational Hierarchy

	SB 1932	WISC 1948	WISC-R 1972	SB 1985	WISC-III 1989	Pooled 1989
Top third	106.28	105.55	104.97	—	105.05	—
Middle third	100.01	99.51	99.55	—	99.91	—
Bottom third	93.71	94.94	95.48	—	95.04	—
IQ gap between top and bottom	12.57	10.61	9.49	8.60[a]	10.01	9.19[a]
Regression correlations[b]	.384	.324	.290	—	.306	—
Correlation ratios[b]	.353	.300	.289	—	.289	—

Note. Trends over time using Stanford–Binet and Wechsler standardization samples, 1932 to 1989, for White children only. Dashes indicate that data were unavailable. SB = Stanford–Binet Intelligence Scale; WISC = Wechsler Intelligence Scale for Children; WISC-R = WISC–Revised; WISC-III = WISC–3rd Edition. For data sources, see Flynn (in press-b), Terman & Merrill (1937, pp. 12, 14, 48; 1973, p. 339), Kaufman & Doppelt (1976), Thorndike, Hagen, & Sattler (1986, pp. 16, 22, 34–35). All WISC data were supplied courtesy of The Psychological Corporation. WISC, Copyright 1949; WISC–R, Copyright 1974; and WISC–III, Copyright 1991, 1974, and 1971, by The Psychological Corporation. Reproduced by permission. All rights reserved. Wechsler Intelligence Scale for Children, WISC, WISC–R, and WISC–III are registered trademarks of The Psychological Corporation.
[a] Data for only White participants from the 1985 SB sample were unavailable; therefore, the class IQ gap for Whites was estimated by deducting the influence of non-White participants as revealed by Wechsler samples. For participants reaching the same age in 1989, the SB data and the WISC-III data were merged to produce the estimate labeled "Pooled 1989." [b] The correlations are between parents' occupational status and children's IQ.

1989. The comparability of the most recent data rests on an assumption: that women show no less merit in attaining professional status than men. Social scientists who find life too dull or devoid of controversy are invited to step forward.

As Table 1 shows, the evidence falsifies the posited trend toward meritocracy. The correlation between children's IQ and their parents' occupational status has been surprisingly stable from 1948 to the present. The pattern is a mean IQ of 105 for upper-class children, 100 for middle-class children, and 95 for lower-class children. The most parsimonious conclusion is this: Nothing, nothing, absolutely nothing has happened. However, the best that evidence can do is show that meritocratic trends do not exist at a particular time and place. This leaves the central contention of the meritocracy thesis untouched. That contention is that if the humane-egalitarian quest of abolishing inequality and privilege is successful, it will result in class stratification of genes for talent, of which IQ is a marker. If such stratification has not occurred, the quest has simply been unsuccessful. Moreover, Herrnstein and Murray (1994) have claimed that a meritocratic future is inevitable. This means that the humane-egalitarian ideal has been given a reprieve both temporary and humiliating. It is a poor ideal that must pray for eternal failure to avoid unwelcome consequences. Therefore, it is necessary to go beyond evidence to analysis.

The case against meritocracy can be put psychologically: (a) The abolition of materialist–elitist values is a prerequisite for the abolition of inequality and privilege; (b) the persistence of materialist–elitist values is a prerequisite for class stratification based on wealth and status; (c) therefore, a class-stratified meritocracy is impossible.

To defend the first proposition, the major barrier to abolition of inequality and privilege is greed and status seeking. Progressive taxation, redistribution of wealth, death duties, welfare, public job creation, publicly funded health care and education, all founder on the rocks of the love of money in one's own pocket, the lust for status superior to one's fellows, the desire to confer advantage for these things on one's family. Test scores and credentials may play a role in the status game, but this does not prevent upper-class parents from giving their children enormous advantages. Educationally efficient homes point them toward superior credentials. Personal networks of relatives, friends, and neighbors locate desirable jobs. The National Center for Career Strategies is reported as stating in 1990 that over 80% of executives find their jobs through networking and that about 86% of available jobs do not appear in the classified advertisements (Ezorsky, 1991, pp. 14–16). Upper-class parents will always find ways of bending the rules in favor of their children. An America in which everyone wants to win the glittering prizes of wealth and status will not pay onerous taxes or show heroic virtue when tempted to seek special advantage, just so the competition can enjoy a level playing field.

To defend the second proposition, were people to lose their obsession with money and status, the class hierarchy that ranks by income and an agreed pecking order of occupations would be diluted beyond recognition. People must *care* about that hierarchy for it to be socially significant or even for it to exist. Imagine a society in which the appreciation of beauty, the pursuit of truth, craft skills, being fit, companionship, family feeling, and so forth really counted for more than having above average income and possessions. Some people would be better than others at all of these things, but there would be at least a score of noncomparable hierarchies, and being better would not necessarily carry financial rewards. Even today, there are executives who care less about promotion than about run-

ning a good 10-kilometer race. The decline of elitist values, less joy in the sheer fact that you are better at something than others, is also relevant. Superior performance would persist, but less status, less passion, less of a sense of being a better human being would attend superior performance. In summary, the meritocracy thesis is psychologically incoherent: It posits a social hierarchy of people who value money and status above all else; and those same people are to ensure that money and status count for nothing in attaining places in that hierarchy.

The case against meritocracy can also be put sociologically: (a) Allocating rewards irrespective of merit is a prerequisite for meritocracy, otherwise environments cannot be equalized; (b) allocating rewards according to merit is a prerequisite for meritocracy, otherwise people cannot be stratified by wealth and status; (c) therefore, a class-stratified meritocracy is impossible.

This reveals an ambiguity at the heart of the meritocracy thesis, namely, failure to specify the quality of the equalized environments assumed. For most of us, giving everyone an equal chance would mean that the *lowest* level of environmental quality would have to be rather good. Yet, equalization is to coexist with a large immiserated underclass, and that class must compete with an elite with an environment so potent that they constitute a menace to democracy (Herrnstein & Murray, 1994, pp. 509–526). The ideal that truly self-destructs in practice is the meritocratic ideal. Those who think it inevitable should give it a plausible social dynamic. They can begin by telling us how environmental equality is to be achieved when a large underclass is already knocking at the door, or conversely, how an underclass is to emerge, if we keep topping up their environmental quality to maintain the level needed for equal opportunity. It is significant that Herrnstein and Murray (1994) imagine environments being equalized by magic. Magic's next task is to reconcile equality with a large underclass. Its final task should be to square the circle.

The sociological analysis reinforces the psychological analysis. The higher the quality of the environment that all enjoy irrespective of merit, the less attractive the prizes left for the winners. If all have decent work, housing, education, health care, security in old age, what remains is not essential for happiness. Many people of talent may want more than the not-unattractive minimum, but how many will care about shaking the last dollar out of the money tree? Social scientists can go on publishing hierarchies that rank the whole population by occupational status, but these will fall short of ranking people by merit, much less by genes for talent. An overenthusiastic sports master can force everyone to participate in the annual school run, but he or she cannot force them to train or try. The published results will not stratify people for genes for running ability.

Analysis of the meritocracy thesis provides not only a rebuttal but also a better understanding of the dynamics of humane-egalitarian ideals. The truth is that we cannot push equality much beyond our capacity to humanize. Every significant step toward equality must be accompanied by an evolution of values unfriendly to success as defined by the present class structure. Humane-egalitarian ideals possess a great glory: a self-correcting mechanism that avoids meritocratic excess. Whatever dark spirits lurk in the depths of equality, meritocracy is not among them.

The Status of Humane Ideals

What summary judgment can we pass about the status of humane-egalitarian ideals? This question leads to two others. Given the present state of our knowledge about group differences, must those ideals be altered to accommodate some unwelcome truths, and must the surgery be so radical as to be life threatening?

I believe that the present state of the evidence is friendly. However, no one can anticipate future evidence, so let us imagine a worst-case scenario: IQ differences must be unambiguously identified with real-world intelligence differences; equalized environments would leave Black Americans with a 10-point IQ deficit thanks to a genetic gap between the races; mean IQ pretty much freezes a group in its place in America's occupational and income hierarchy. Clearly, these theses would be unwelcome. They would mean that there was no hope of things highly desirable, that all ethnic groups would be equally safe from cycles of boom and bust and economic deprivation, that all ethnic groups would make contributions to the intellectual and scientific life of America that would command roughly equal respect.

They would not mean, as already indicated, that there was no case for affirmative action. They would not mean abandoning the ideal of giving every American a decent quality of life, and if that were done, the powerful emotions engendered by group differences in test scores, academic achievement, occupation, and income would be moderated. The intensity of those emotions is the product of human misery. It is one of the great strengths of humane ideals that they can adapt better than most ideals when confronted with unwelcome truths. A classical racist believes that Black is always worse than White. That belief cannot be defended no matter whether Jensen is correct or I am correct: Even half of Blacks overlapping with the top 84% of Whites for desirable personal traits is enough to deliver it a mortal blow. Humane ideals can survive because their champions can still ask how can the handicaps of group membership be minimized, how can inequality be reduced to acceptable levels, how can privilege be curtailed (Flynn, 1998b)?

However, the adaptation of humane-egalitarian ideals would not be without pain, and therefore, the evidence for unwelcome theses should be convincing. Galileo is so often championed against Cardinal Bellarmine that equity demands a word or two in defence of the latter. The evidence about whether the sun is at the center of the solar system was mixed. The predicted stellar parallax was not there; Galileo needed more epicycles than the old astronomy did; he had the earth circling not the sun but a point in space somewhat removed from the sun! Bellarmine did not deny that the thesis might be true; he did not deny that it should be tested against evidence. He argued that it should not be *asserted* to be true until the evidence was decisive, given

that it required a reinterpretation of scripture unsettling to those of simple faith. Bellarmine likened the Devil to a wily angler and people to frogs with mouths gaping, ever ready to take the Devil's bait. Humane ideals can adapt to significant genetic differences between the races, but there are a lot of frogs out there with mouths gaping for the bait of racism. Those who believe in such differences should be very, very certain before they stamp them as proven. Needless to say, those skeptical of such differences should not disgrace themselves: They must not follow Bellarmine so far as to try to silence opponents whose confidence, arguably at least, outruns the evidence. The indefensible suppression of Chris Brand's book is a case in point.

The meritocracy thesis is another matter. It is life threatening because humane-egalitarian ideals cannot survive the claim that they self-destruct in practice. Charitable feelings toward others could survive, but not the great ideal of reducing inequalities of environment and privilege, not if every step in that direction is a step toward an even more obnoxious inequality. If the refutation of the meritocracy thesis offered herein is found wanting, other humanistic scholars had better do some serious thinking. This sombre note concludes the 20-year holiday of a philosopher visiting another discipline. The dark side has been a perception of what is at stake. The delight has been the interplay of ideas with a scholar of the learning and subtlety of Arthur Jensen.

REFERENCES

Backman, M. E. (1972). Patterns of mental abilities: Ethnic, socioeconomic and sex differences. *American Educational Research Journal, 9,* 1–12.

Bouchard, T. J., Jr., & McGue, M. (1981). Familial studies of intelligence: A review. *Science, 212,* 1055–1059.

Brand, C. R. (1987a). Bryter still and more bryter? *Nature, 328,* 110.

Brand, C. R. (1987b). Keeping up with the times. *Nature, 328,* 761.

Coleman, J. (1966). *Equality of educational opportunity* (supplemental appendix, section 9.10, correlation tables). Washington, DC: U.S. Office of Education.

Eysenck, H. J. (1981). Special review: James R. Flynn, *Race, IQ and Jensen. Personality and Individual Differences, 2,* 259.

Ezorsky, G. (1991). *Racism and justice: The case for affirmative action.* Ithaca, NY: Cornell University Press.

Floud, R., Wachter, K., & Gregory, A. (1990). *Height, health and history.* Cambridge, England: Cambridge University Press.

Flynn, J. R. (1967). *American politics: A radical view.* Auckland, New Zealand: Blackwood & Janet Paul.

Flynn, J. R. (1980). *Race, IQ and Jensen.* London: Routledge.

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

Flynn, J. R. (1985). Wechsler intelligence tests: Do we really have a criterion of mental retardation? *American Journal of Mental Deficiency, 90,* 236–244.

Flynn, J. R. (1987a). Flynn replies to Nichols. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 234–235). Lewes, England: Falmer Press.

Flynn, J. R. (1987b). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 171–191.

Flynn, J. R. (1987c). The ontology of intelligence. In J. Forge (Ed.), *Measurement, realism and objectivity* (pp. 1–40). Dordrecht, the Netherlands: Reidel.

Flynn, J. R. (1987d). Race and IQ: Jensen's case refuted. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 221–232). Lewes, England: Falmer Press.

Flynn, J. R. (1989). Rushton, evolution, and race: An essay on intelligence and virtue. *The Psychologist, 2,* 363–366.

Flynn, J. R. (1991a). *Asian Americans: Achievement beyond IQ.* Hillsdale, NJ: Erlbaum.

Flynn, J. R. (1991b). Reaction times show that both Chinese and British children are more intelligent than one another. *Perceptual and Motor Skills, 72,* 544–546.

Flynn, J. R. (1992). Cultural distance and the limitations of IQ. In J. Lynch, C. Modgil, & S. Modgil (Eds.), *Education for cultural diversity: Convergence and divergence* (pp. 343–360). London: Falmer Press.

Flynn, J. R. (1993). Skodak and Skeels: The inflated mother–child IQ gap. *Intelligence, 17,* 557–661.

Flynn, J. R. (1994a). Giving *g* a fair chance: How to define intelligence, survive falsification, and resist behaviorism. *Psychological Inquiry, 5,* 202–208.

Flynn, J. R. (1994b). IQ gains over time. In R. J. Sternberg (Ed.), *The encyclopedia of human intelligence* (pp. 617–623). New York: Macmillan.

Flynn, J. R. (1996). Group differences: Is the good society impossible? *Journal of Biosocial Science, 28,* 573–585.

Flynn, J. R. (1998a). Evidence against Rushton: The genetic loading of WISC-R subtests and the causes of between-group IQ differences. *Personality and Individual Differences, 26,* 373–379.

Flynn, J. R. (1998b). *How to defend humane ideals: Substitutes for objectivity.* Manuscript submitted for publication.

Flynn, J. R. (1998c). IQ gains over time: Toward finding the causes. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 25–66). Washington, DC: American Psychological Association.

Flynn, J. R. (1998d). Israeli military IQ tests: Gender differences small; IQ gains large. *Journal of Biosocial Science, 30,* 541–553.

Flynn, J. R. (1998e). *New data and hypotheses: Massive gains on Similarities; IQ gains equal fluid-g gains; concerning the cause of IQ gains; the Spearman hypothesis irrelevant to race and IQ.* Manuscript submitted for publication.

Flynn, J. R. (1998f). Reply to Rushton: A gang of *gs* overpowers factor analysis. *Personality and Individual Differences, 26,* 391–393.

Flynn, J. R. (1998g). The schools: IQ tests, labels, and the word "intelligence." In J. S. Carlson, J. Kingma, & W. Tomic (Eds.), *Advances in cognition and educational practice: Vol. 5. Conceptual issues in research on intelligence* (pp. 13–42). London: JAI Press.

Flynn, J. R. (1998h). WAIS-III and WISC-III: IQ gains in the United States from 1972 to 1995; how to compensate for obsolete norms. *Perceptual and Motor Skills, 86,* 1231–1239.

Flynn, J. R. (in press-a). The hidden history of IQ and special education: Can the problems be solved? *Psychology, Public Policy, and Law.*

Flynn, J. R. (in press-b). IQ trends over time: Intelligence, race, and meritocracy. In S. Durlauf, K. Arrow, & S. Bowles (Eds.), *Meritocracy and inequality.* Princeton, NJ: Princeton University Press.

Gottfredson, L. S. (1987). The practical significance of Black–White differences in intelligence. *Behavioral and Brain Science, 10,* 510–512.

Greenfield, P. (1998). The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 81–123). Washington, DC: American Psychological Association.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class in American life.* New York: Free Press.

Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39,* 1–123.

Jensen, A. R. (1972). *Genetics and education.* New York: Harper & Row.

Jensen, A. R. (1973a). *Educability and group differences.* New York: Harper & Row.

Jensen, A. R. (1973b). *Educational differences.* London: Methuen.

Jensen, A. R. (1980). *Bias in mental testing.* London: Methuen.

Jensen, A. R. (1981). *Straight talk about mental tests.* New York: Free Press.

Jensen, A. R. (1987a). Differential psychology: Towards consensus. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 353–399). Lewes, England: Falmer Press.

Jensen, A. R. (1987b). The *g* beyond factor analysis. In R. R. Ronning, J. A. Glover, J. C. Conoley, & J. C. Witt (Eds.), *The influence of cognitive psychology on testing* (pp. 87–142). Hillsdale, NJ: Erlbaum.

Jensen, A. R. (1988). *The Oddman-Out Test: A chronomatic anchor for psychometric tests of g.* Unpublished manuscript.
Jensen, A. R. (1989). Raising IQ without increasing *g?* A review of *The Milwaukee Project: Preventing mental retardation in children at risk. Developmental Review, 9,* 234–258.
Jensen, A. R. (1992). Commentary: Vehicles of *g. Psychological Science, 3,* 275–278.
Jensen, A. R. (1997). Psychometric *g* and the race question. In J. S. Carlson, J. Kingma, & W. Tomic (Eds.), *Advances in cognition and educational practice: Vol. 4. Reflections on the concept of intelligence* (pp. 1–23). London: JAI Press.
Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.
Kaufman, A. S., & Doppelt, J. E. (1976). Analysis of WISC-R standardization data in terms of stratification variables. *Child Development, 47,* 165–171.
Lewontin, R. C. (1976a). Further remarks on race and the genetics of intelligence. In N. J. Block & G. Dworkin (Eds.), *The IQ controversy* (pp. 107–112). New York: Pantheon Books.
Lewontin, R. C. (1976b). Race and intelligence. In N. J. Block & G. Dworkin (Eds.), *The IQ controversy* (pp. 78–92). New York: Pantheon Books.
Loehlin, J. C. (1989). Partitioning environmental and genetic contributions to behavioral development. *American Psychologist, 44,* 1285–1292.
Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence.* San Francisco: Freeman.
Loehlin, J. C., Vandenberg, S. C., & Osborne, R. T. (1973). Blood group genes and Negro–White ability differences. *Behavior Genetics, 3,* 263–270.
Lynn, R. (1987). Japan: Land of the rising IQ. A reply to Flynn. *Bulletin of the British Psychological Society, 40,* 464–468.
Lynn, R. (1989). Positive correlation between height, head size and IQ: A nutrition theory of the secular increases in intelligence. *British Journal of Educational Psychology, 59,* 372–377.
Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences, 17,* 257–271.
Lynn, R. (1996). *Dysgenics: Genetic deterioration in modern populations.* Westport, CT: Praeger.
Mackenzie, B. (1984). Explaining race differences in IQ. *American Psychologist, 39,* 1214–1233.
Neisser, U. (1998). Introduction: Rising test scores and what they mean. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 3–22). Washington, DC: American Psychological Association.
Nichols, R. C. (1987). Racial differences in intelligence. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 213–220). Lewes, England: Falmer Press.
Plomin, R., & Rende, R. (1991). Human behavioral genetics. *Annual Review of Psychology, 42,* 161–190.
Raven, J. (1995). Methodological problems with the 1992 standardization of the SPM: A response. *Personality and Individual Differences, 18,* 443–445.
Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales* (Section 1). Oxford, England: Oxford Psychologists Press.
Rushton, J. P. (1997). Race, intelligence, and the brain: The errors and omissions of the "revised" edition of S. J. Gould's *The Mismeasure of man* (1996). *Personality and Individual Differences, 23,* 169–180.
Rushton, J. P. (1998). 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,* 381–389.
Scarr, S., Pakstis, A. J., Katy, S. H., & Barber, W. B. (1977). Absence of a relationship between degree of White ancestry and intellectual skills within a Black population. *Human genetics, 39,* 69–86.
Scarr, S., Weinberg, R. A., & Waldman, I. D. (1993). IQ correlations in transracial adoptive families. *Intelligence, 17,* 541–555.
Schoenthaler, S. J., Amos, S. P., Eysenck, H. J., Peritz, E., & Yudkin, J. (1991). Controlled trial of vitamin–mineral supplementation: Effects on intelligence and performance. *Personality and Individual Differences, 12,* 351–362.
Schooler, C. (1998). Environmental complexity and the Flynn effect. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 67–79). Washington, DC: American Psychological Association.
Sowell, T. (1972). *Black education: Myths and tragedies.* New York: McKay.
Stevenson, H. W., Stigler, J. W., Lee, S., Lucker, G. W., Kitamura, S., & Hsu, C. (1985). Cognitive performance and academic achievement of Japanese, Chinese, and American children. *Child Development, 56,* 718–734.
Storfer, M. D. (1990). *Intelligence and giftedness: The contributions of heredity and early environment.* San Francisco: Jossey-Bass.
Terman, L. M., & Merrill, M. A. (1937). *Measuring intelligence.* London: Harrap.
Terman, L. M., & Merrill, M. A. (1973). *Stanford–Binet Intelligence Scale 1973 norms edition.* Boston: Houghton Mifflin.
Thorndike, R. L., Hagen, E. P., & Sattler, R. L. (1986). *Technical manual: Stanford–Binet Intelligence Scale fourth edition.* Chicago: Riverside.
Tizard, B. (1974). IQ and race. *Nature, 247,* 316.
Tizard, B., Cooperman, O., Joseph, A., & Tizard, J. (1972). Environmental effects on language development: A study of young children in long-stay residential nurseries. *Child Development, 43,* 337–358.
Vernon, P. E. (1982). *The abilities and achievements of Orientals in North America.* New York: Academic Press.
Vincent, K. R. (1993). On the perfectibility of the human species: Evidence using fixed reference groups. *Texas Counselling Association Journal, 22,* 60–64.
Wechsler, D. (1992). *WISC-III manual: Australian adaptation.* San Antonio, TX: Psychological Corporation.
Williams, W. (1998). Are we raising smarter kids today? School- and home-related influences on IQ. In U. Neisser (Ed.), *The rising curve: Long-term gains in IQ and related measures* (pp. 125–154). Washington, DC: American Psychological Association.
Witty, P. A., & Jenkins, M. D. (1934). The educational achievements of a group of gifted Negro children. *Journal of Educational Psychology, 25,* 585–597.
Witty, P. A., & Jenkins, M. D. (1936). Inter-race testing and Negro intelligence. *Journal of Psychology, 1*(2), 179–192.
Yerkes, R. M. (1921). Psychological examining in the United States Army. In *Memoirs of the National Academy of Sciences* (Vol. 15). Washington, DC: U.S. Government Printing Office.