The Rising Curve

Long-Term Gains in IQ and Related Measures

Edited by Ulric Neisser

American Psychological Association • Washington, DC

Contents

Contributors

Mark Berends, RAND Institute for Education and Training, Washington, DC

Stephen J. Ceci, Department of Human Development, Cornell University

James R. Flynn, Political Studies, University of Otago, Dunedin New Zealand

Patricia M. Greenfield, Department of Psychology, University o California–Los Angeles

David W. Grissmer, RAND Institute for Education and Trainin Washington, DC

Robert M. Hauser, Center for Demography, University of Wisc

Min-Hsiung Huang, Institute of European and American Studie Academia Sinica, Taipei, Taiwan

Sheila Nataraj Kirby, RAND Institute for Education and Trainir Washington, DC

Matthew Kumpf, Department of Human Development, Cornell University

John C. Loehlin, Department of Psychology, University of Texas

Richard Lynn, University of Ulster at Coleraine, Northern Irelan

Reynaldo Martorell, Rollins School of Public Health, Emory University

Ulric Neisser, Department of Psychology, Cornell University

Samuel H. Preston, Population Studies Center, University of Pennsylvania

Tina B. Rosenblum, Department of Human Development, Corne University

Carmi Schooler, Laboratory of Environmental Studies, National Institute of Mental Health, Rockville, MD

1

Introduction: Rising Test Scores and What They Mean

Ulric Neisser

For the better part of a century, Americans have been giving intelligence and achievement tests and viewing the results with alarm. As early as the 1920s, for example, the country was dismayed to learn that the average American man had the mental age of a 14-year-old (Lippmann, 1976). In the 1970s, Scholastic Aptitude Test (SAT) scores seemed to be in free fall; in the 1980s, children were found to lag behind their counterparts in much of the world in their knowledge of science and math. Throughout this time, the average IQ and school achievement scores of Black Americans remained substantially below those of Whites—a gap so persistent that some theorists came to regard it as inevitable. In such a context, the findings reported in this book come as a welcome surprise. Scores on intelligence tests are *rising*, not falling; indeed, they have been going up steeply for years. This rapid rise is not confined to the United States; comparable gains have occurred all over the industrialized world. A second major finding—that the gap in school achievement between Black and White children has diminished sharply in recent years—is at least equally important. These remarkable, environmentally driven trends deserve our most thoughtful attention.

Psychometrics, the study of mental tests, is a rather esoteric busi-

ness. From time to time, however, new findings or new claims about test scores seem to erupt into public consciousness. The most recent of those eruptions occurred in 1994, when Richard Herrnstein and Charles Murray published a best-selling book called *The Bell Curve*. Their analysis was profoundly pessimistic. Like many psychometricians, they began by assuming that scores on intelligence tests chiefly reflect a single underlying ability, called *g*. Some people have more *g* and others less, depending in large part on their genetic endowment. According to Herrnstein and Murray, an individual's *g* largely determines what he or she can achieve in today's complex society. They also thought it likely that genetic factors contribute to the difference between the average IQs of Whites and Blacks, a gap which shows few signs of narrowing. As if all that were not discouraging enough, *The Bell Curve* also devoted a chapter to so-called *dysgenic trends*: If *g* is highly heritable and if low-*g* people consistently have more children than high-*g* people do, is not the overall intelligence of the population bound to decline in the long run?

This book has a very different take on all three facets of that argument: on what intelligence tests measure, on the difference between the test scores of Black and White Americans, and on the likelihood that there are long-term dysgenic trends. Our starting point is the surprising and continuing rise in test scores: Performance on broad-spectrum tests of intelligence has been going up about 3 IQ points per decade ever since testing began. (The practice of restandardizing the major tests from time to time has kept the mean IQ at about 100 despite these gains, making these changes harder to see.) Even more surprising is that scores on specialized tests of abstract reasoning like the Raven Progressive Matrices—often described as the very best measure of *g*—are rising still faster. Herrnstein and Murray were aware of these gains but gave them short shrift—an understandable decision, considering how profoundly they undermine many of the claims of *The Bell Curve*.

IMPLICATIONS OF THE GAINS

Psychometricians have long known that test performance tends to rise from one generation to the next. The average scores of American draftees in the second world war, for example, were far higher than in the first world war (Tuddenham, 1948). Students of adult intellectual development have also noted the existence of generational differences (Schaie, 1983, 1997). Nevertheless, the size and significance of the gains were not widely appreciated until they were systematically documented by James Flynn, a political scientist at the University of Otago in New Zealand (Flynn, 1984, 1987). Herrnstein and Murray christened them the "Flynn effect," and the name has stuck. No matter how the Flynn effect is eventually explained, it presents grave difficulties for all three facets of the pessimistic argument in *The Bell Curve*. Let us consider them one at a time:

1. Is there a single underlying *g*, largely determined by genetic factors? If this were true, there would be only two possibilities: The Flynn rise either does or does not reflect real increases in *g*. If it does reflect real increases, *g* clearly is affected by environmental factors because no genetic process could produce such large changes so quickly. Whatever those environmental factors may be, we can at least reject the hypothesis that intelligence is genetically fixed. But if it does not reflect real increases—if, as Flynn himself believes, the gains only reflect some trivial artifact—then the tests are evidently flawed, and all arguments based on test scores become suspect. Either way, things look bad for *g* and the arguments of *The Bell Curve*.
2. Does the mean Black–White difference on IQ tests reflect a genetic—or at least very firmly entrenched—limitation on the mental abilities of average Black Americans? In the 1930s, the average Black–White IQ difference was about 15 points. Half a century later in the 1980s, it was still about 15 points. Do these findings imply that nothing much had changed? On the contrary, given the Flynn rise, *both* groups gained some 15 points during those 50 years! The gains made by Blacks are especially impressive because they closed what was once the entire gap: Blacks in the 1980s performed at the level of Whites in the 1930s. Perhaps it was once hard to believe that environmental factors could sustain (or elimi-

nate) a difference of this magnitude, but in the light of these gains it is not hard to believe today. In addition there are new findings, based on school achievement tests rather than intelligence tests, which show substantial reductions in an analogous Black–White gap. Part 2 of this volume reviews those findings and their significance.

3. Is the overall genetic potential of the population bound to decline because people of low IQ have more children than people of high IQ? This is an old bogey. First conceived by Francis Galton in the 19th century, it continued to worry his successors in the "eugenics" movement for decades. Flynn's findings suggest that the concern was misplaced: Intelligence has in fact been rising, not falling. Nevertheless, the issue is not quite closed. Because people of low IQ *do* tend to have more children, scholars who worry about eugenic issues can argue that the decline *must* be occurring. Although that (hypothetical) downward trend is currently masked by an environmentally driven rise, it might still be something to worry about. In Part 3 of this book, Richard Lynn insists that the danger is real, whereas a number of critics argue to the contrary. One of those critics, Samuel Preston, shows mathematically that Galton's basic assumption was simply wrong: Differential birth rates do *not* necessarily produce changes in a population mean.

One way or another, all these issues revolve around test scores. For that reason, it is important to say a few words about the tests themselves and the concepts of intelligence that they help to support. There have been two major strands in the history of psychometrics, both now about a century old. One strand, which can be traced back to the British psychologist Charles Spearman (1904), has focused on the theoretical analysis of relations among tests; the concept of *g* arose in this tradition. The other strand began with Alfred Binet in France but is now primarily an American enterprise; it has been more closely driven by practical considerations. The latter tradition had given us the concept of IQ as well as the distinction between tests of intelligence and of achievement.

FACTOR ANALYSIS AND *g*

Theoretical analysis in the first of these traditions begins with correlations. Suppose that two different tests have been administered to the same group of individuals; did people who got high scores on A also tend to get high scores on B? To get a quantitative answer to this question, one uses the paired test scores to calculate the correlation coefficient *r*. Any substantial positive value of *r* indicates that people who scored high on A tend to score high on B as well. If *r* reaches its maximum value of 1.00, A and B are very closely related indeed: Scores on either one of the tests can be perfectly predicted by scores on the other. A negative value of *r* (not likely with tests of mental ability) implies that high scores on A go with *low* scores on B and vice versa. An *r* near zero means that the tests are essentially independent.

When more than two tests are involved, the analysis gets more complicated. Three tests generate three different values of *r* (A with B, A with C, B with C), and the pattern of those correlations may be interesting. If r_{AB} is very high and r_{AC} and r_{BC} are both near zero, for example, we might conclude that Tests A and B measure roughly the same thing, whereas Test C measures something different. This would amount to saying that there are only two underlying *factors* even though three tests were given. A larger battery of tests would generate too many *r*s for this kind of intuitive reasoning, but one can still arrange them in a correlation matrix like that in Table 1. The data illustrated there are typical: The *r*s are all positive but well below 1.00. It is an empirical fact that almost every battery of mental tests produces just such a *positive manifold* (at least if the sample of test takers is representative of the general population). This means that people who do well on one test tend also to do well on others, in varying degrees that depend on the tests in question. Why should this be true?

Spearman's (1904) approach to this problem was to invent a new statistical method, one that is still widely used today. The purpose of *factor analysis* is to determine how well the data in any given correlation matrix can be fit by a model based on a smaller number of underlying factors. Spearman's own analyses convinced him that only a single factor—he called it *g* for "general ability"—was needed to account for

Table 1

Intercorrelations of Six U.S. Navy Classification Tests

Test	Reading	Arithmetic Reasoning	Mechanical Aptitude	Electrical Knowledge	Mechanical Knowledge
General Classification	.81	.69	.60	.53	.49
Reading		.69	.56	.51	.46
Arithmetic Reasoning			.61	.47	.41
Mechanical Aptitude				.53	.55
Electrical Knowledge					.78
Mechanical Knowledge					

Note. An illustration of the positive manifold. From *Essentials of Psychological Testing* (p. 312, Table 10.3), by L. J. Cronbach, 1970, New York: Harper & Row. Copyright 1970 by HarperCollins. Reprinted with permission.

the positive correlations in the matrix. Individuals differed in how much *g* they had, and *g* contributed in varying degrees to performance on different tests (each of which also measured a more specific ability). Individual differences in *g* itself were most important for tests involving some form of abstract reasoning, so Spearman concluded that *g* represents a core ability to extract (or "educe") logical relations. He and his successors regarded *g* as the central essence of intelligence.

The hypothesis that there is a single basic form of intelligence—and that some of us have more of it than others—seems to fit a lot of people's intuitions. (As we have seen, it is central to the argument of *The Bell Curve.*) Nevertheless, it is by no means a necessary conclusion. Many different forms of factor analysis have been devised over the years, and they do not all find the same structure; in particular, they do not all find a single *g*. Some analysts have identified a half-dozen partly independent forms of intelligence, while others have argued that all such forms still incorporate something like *g*. There are theorists who regard *g* and all other products of factor analysis as little more than statistical artifacts; on the other side are many who believe, even more

strongly than Spearman, that it reflects an inherited and very basic property of the brain.

THE RAVEN

Whatever *g* may be, we at least know how to measure it. The accepted best measure, which has played a central role in analyses of the worldwide rise in test scores, is the Raven Progressive Matrices. This test, devised by Spearman's student John C. Raven, was first published in 1938 and is now available at several levels of difficulty. Arthur Jensen has said that Raven's test "apparently measures *g* and little else" (1980, p. 646) and that it "is probably the surest instrument we now possess for discovering intellectually gifted children from disadvantaged backgrounds" (p. 648).

Raven's test consists of a series of items whose difficulty varies systematically. Each item consists of a 3 × 3 matrix with one empty cell; the test taker must decide which of eight candidate entries (shown below the matrix) would best complete it. Figure 1 shows a relatively difficult Raven-type item. One way to solve this particular item is to discover that the following principle applies to the first two rows (as well as to the first two columns): The entry in the third cell results from superimposing those in the first two cells, while deleting all line segments that they have in common. Only one of the candidate entries (No. 5) would complete the matrix in a way that fits this principle.

The Raven is of particular interest because it shows such large IQ gains over time. In The Netherlands, for example, all male 18-year-olds take a version of the Raven as part of a military induction requirement. The mean scores of those annual samples rose steadily between 1952 and 1982, gaining the equivalent of 21 IQ points in only 30 years! This amounts to a rate of no less than 7 points per decade—a figure confirmed by data from many other countries (Flynn, 1987). What can these increases mean?

One way to address this question is to look more closely at the abilities that the Raven requires. The example in Figure 1 makes some of them obvious: One must be able to "educe" abstract relations among

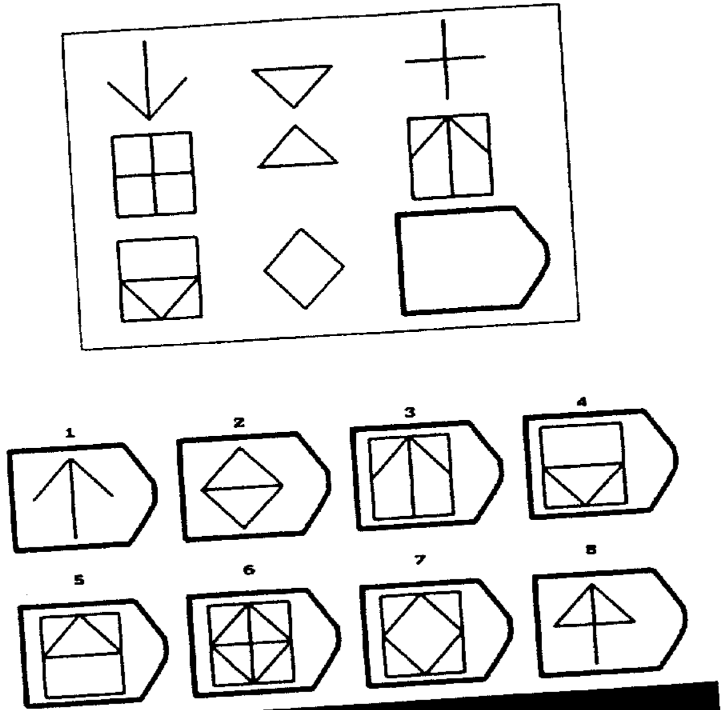

Figure 1

A problem similar to those in the Raven Advanced Progressive Matrices. The matrix at the top of the figure is missing one element; which of the eight pieces at the bottom would complete it appropriately? From "What One Intelligence Test Measures," by P. A. Carpenter, M. A. Just, and P. Shell, 1990, *Psychological Review, 97,* p. 409, Figure 4C. Copyright 1990 by the American Psychological Association. Reprinted with permission.

a series of meaningless figures and to keep track of several such series at once (Carpenter, Just, & Shell, 1990). A special form of visual analysis is also required: Each entry must be dissected into the simple line segments of which it is composed before the processes of abstraction can operate. Although this kind of seeing is rarely needed in the ordinary environment (where figures are usually seen as wholes), it may have

become more familiar in recent years. As our exposure to movies, television, and other technical optical displays increases, our skills of visual analysis may be increasing too. In chapter 4 of this volume, Patricia Greenfield explores the hypothesis that those very changes are responsible for the Flynn rise.

STANDARD TESTS OF "IQ"

The other main strand in the history of testing begins with a more practical problem. Some children do not do well in school; how can they be helped? According to Alfred Binet, who addressed this question in the early 1900s, a good first step is to determine the child's level of intellectual maturity, that is, his or her mental age (MA). No theory of intelligence is necessary for this purpose; one only needs a pool of test items that have already been presented to samples of children of different ages. Binet's samples showed, for example, that the average French 5-year-old could copy a square, count four pennies, indicate the heavier of two cubes, and so on; 7-year-olds could copy a diamond and define familiar objects in terms of their use; 9-year-olds knew the day of the week and could make change in simple play-store transactions. A child's mental age can be established simply by seeing where he or she succeeds and begins to fail in such an age-graded series of items.

Binet was trying to help children learn, not to assign a fixed level of intelligence to any given child. His tests were so practical and their results so consistent, however, that others were soon using them for that very purpose. One need only divide a child's MA by his or her actual age (and multiply the result by 100) to get an intelligence quotient (IQ), which has useful metric properties. IQs predict school grades rather well (although far from perfectly); they are fairly stable throughout the developmental years (although large changes sometimes occur); and their distribution in a given age group roughly follows the bell-shaped normal curve (not quite, but I ignore the discrepancies here). IQ testing quickly became popular, especially in the United States, where Lewis M. Terman soon standardized a test based on Binet's principles. He called it the Stanford–Binet.

Since that time, many different kinds of intelligence tests have come on the market. In a particularly influential series developed by David Wechsler—including the Wechsler Intelligence Scale for Children (WISC) and the Wechsler Adult Intelligence Scale (WAIS)—each test is organized into a dozen subscales that each include a single type of item: vocabulary, comprehension, block design, and so on. In modern tests like these, IQs are defined in ways that do not involve quotients. Nevertheless, the definition always begins with the scores of a *standardization sample*, selected to represent various population age groups, that took the test at a certain point in time. The average score for a given age group in that sample defines an IQ of 100; its standard deviation (a statistical measure of spread around the mean) defines 15 IQ points.

IQ GAINS OVER TIME

At first glance it would seem that by this definition, the mean IQ must always be 100. Moreover, given the properties of the normal curve, about 2 ½% of the population should always have IQ scores above 130 (such individuals are called "very superior"), and another 2 ½% should score below 70 (called "intellectually deficient"). However, these implications hold only as long as the standardization sample represents the current population, which may not be the case if that population's competence has changed. It is partly to guard against this possibility that tests are restandardized from time to time. The WAIS, for example, was originally normed in 1953; its successor, the WAIS-R, was normed on a new sample in 1978. Wechsler (1981) then asked a group of people to take both tests; the result was that they scored 7.5 points higher on the (older) WAIS! Given that 25 years had elapsed between standardizations, this represents a gain of 3 points per decade.

Wechsler's (1981) result was no fluke. In almost every study in which the same group has taken two tests standardized at different times, scores have been higher on the earlier test (Flynn, 1984). Generally speaking, a person who has a given IQ score with respect to his or her contemporaries scores substantially higher when compared with

earlier samples. These gains are steady and systematic: Performance on IQ tests has been going up 3 points per decade ever since the first Stanford–Binet was normed in 1932. Although this rise is smaller than the remarkable increases on the Raven (see Figure 2), it is still impressive.

The seven chapters in Part 1 of this volume are devoted to these remarkable gains. Flynn himself argues in chapter 2 that they are too large to be genuine; we cannot possibly be that much smarter than our grandparents! His is a minority position. All the other contributors to Part 1 believe that there have been real and substantial gains in at least some kinds of intelligence. But then, what is responsible for them?

A wide range of candidate explanations—sociological, psychological, and biological—is considered in these pages. The sociological analysis is presented by Carmi Schooler (chapter 3), who reviews existing evidence on the effects of environmental complexity and modernization as part of his general critique of Flynn's argument. In chapter 4, Patricia Greenfield focuses on technologically driven changes in the visual environment, changes that may well have had specific effects on the modes of thought required by mental tests. In chapter 5, Wendy Williams reviews a broad spectrum of school- and home-related variables that may have contributed to the rise in different ways. The last three chapters in Part 1 focus on the nutrition hypothesis: Perhaps people have been getting smarter in the last hundred years or so for the same biological reason—dietary improvement—that they have been getting taller! (On average, taller people tend to have larger heads and brains.) Richard Lynn supports this explanation strongly in chapter 8, whereas Reynaldo Martorell is skeptical (chapter 7); Marian Sigman and Shannon Whaley take an intermediate position in chapter 6.

Even this wide-ranging set of hypotheses does not exhaust the possibilities. Robert Zajonc (1976) has long argued that intelligence is affected by family size and sibling position. The relationship is complex, but in most cases early-born children have an advantage over later borns because they spend more time in more intelligent company (that of their parents). Therefore, the long-term demographic trend toward smaller families, which implies that fewer children are being born in

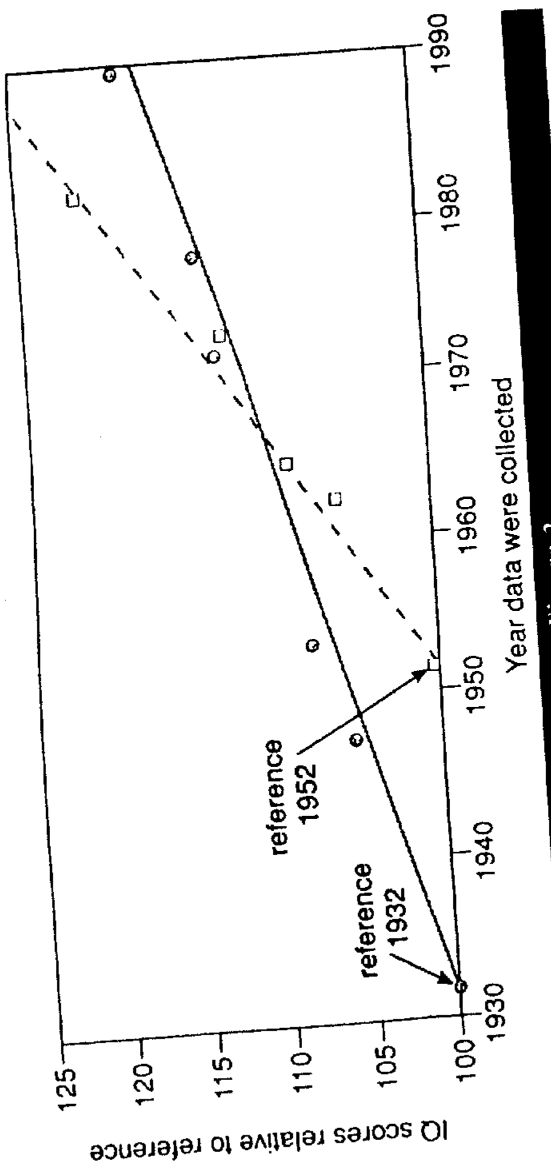

Figure 2

The rise in intelligence test scores since 1932. The solid line shows the adjusted mean performance of Americans on various Wechsler tests as well as the Stanford–Binet; the dashed line shows the mean performance of 18-year-old Dutch males on a version of Raven's Progressive Matrices. Data from Neisser (1997b).

later sibling positions, may have contributed to raising the average level of intelligence (Zajonc & Mullally, 1997).

GENETIC VERSUS ENVIRONMENTAL INFLUENCES

Terman and many of his successors were "hereditarians." They assumed both (a) that people's intellectual abilities—specifically, their IQ scores—are determined at or before birth by innate endowment; and (b) that *differences* in IQ scores of individuals within a given group—say, the adult White American male population of a given year—primarily reflect differences in those endowments. The early hereditarians did not distinguish between these two claims, but the development of modern behavior genetics has made it obvious that they are distinct. The first is false, but the second may not be far from the truth.

Every human trait develops through an interaction between genetic factors and the (internal and external) environment. Indeed, the environment must lie within a certain normal range if the trait is to exist at all. Intelligence is a case in point: No one could acquire cultural information, learn a language, or master any important cognitive skill without environmental support. Although this fact alone is enough to show that IQ scores are not determined exclusively by innate endowment, the existence of the Flynn effect makes that point in a more concrete way. However, one may choose to interpret it, the fact that (unknown) environmental factors are raising the mean IQ of Americans by 3 points per decade certainly shows that the environment matters!

The second proposition has quite a different status. Within a given population and a given range of environments (e.g., those that are characteristic for White American males in 1998), genetic factors do make a major contribution to individual differences. This has now been shown beyond a reasonable doubt by the methods of behavior genetics, a discipline that is primarily concerned with variability. The individuals in a given population differ on almost any measure one is likely to care about: their heights, weights, Raven scores, IQ scores, or anything else. Every such measure has a distribution, often a bell-shaped normal one.

Roughly speaking, there are two reasons for these individual differences. On the one hand, people differ in the sets of genes with which they are born (their *genomes*), and this may affect the trait in question. (Except for identical twins, no two people have the same genome.) On the other hand, people differ in the specific environments that they have encountered. An interesting question, then, is how much of the trait's total variability (technically, its variance) is due to genetic differences and how much to differences in environment.

The proportion of any trait's variance due to genetic differences is called its heritability (h^2). An h^2 of zero means that the genes make no contribution to the variance, whereas an h^2 of 1.00 means that genes account for all of it. In fact, it is hard to think of traits that fit either of those profiles. Most measurable characteristics are somewhat heritable, that is, have intermediate values of h^2. Heritabilities can be estimated on the basis of correlations between relatives: identical twins, fraternal twins, siblings, adopted children and their adoptive or natural parents, and so on. All these estimates have technical complexities that I cannot consider here. (One of the most interesting of those complexities is that some variables seem to straddle the genetic–environmental dichotomy and cannot readily be assigned to either side; an example is given below.) Nevertheless, the overall pattern of results is clear: The intelligence measured by test scores, like most other human traits, is at least moderately heritable.

It is now widely agreed that h^2 for IQ lies between .40 and .80 in the U.S. White population. In other words, genetic factors contribute substantially to individual differences in intelligence. (That is why some theorists, including the authors of *The Bell Curve*, have been so concerned about long-term dysgenic trends. We shall see in Part 3 that their concern may be misplaced.) But by the same token, there is also no doubt that the environment contributes to those differences: h^2 is well below 1.00. Unfortunately, no one knows what it is about the environment that makes this contribution to differences in IQ scores. Some obvious possibilities, such as the economic and intellectual quality of children's home situations, may be less important than was once believed. The surprising fact is that when biologically unrelated children

are raised in the same home (as in many cases of adoption), the correlation between their IQ scores is unimpressive in childhood and near zero as they grow up! This finding is important, but it is still negative: The aspects of the environment that *do* matter for the development of intelligence have not yet been identified.

In fact, there are surprisingly many things about intelligence that no one presently understands. One example has just been noted: There are evidently aspects of the environment that affect individual differences, but we do not know what those aspects are or how they work. Another case in point, considered earlier, is the Flynn effect itself. The worldwide rise in test scores is surely driven by some kinds of environmental change, but what are they? A third example, to be discussed later in this chapter, concerns the difference between the mean IQ scores of Blacks and Whites. As we shall see, its causes are equally mysterious.

One further discovery of modern behavior genetics should be mentioned at this point. Several different observations, including the pattern of correlations between IQs of adopted children mentioned above, have led to an unexpected discovery: Heritability itself seems to increase with age. Genetic differences contribute relatively more to the variability of IQ among *adults*, whereas environmental differences contribute relatively more to its variability among *children*. What can this mean? One possibility is that the environment does matter but that adults are somewhat able to choose their own environments (on the basis of genetically influenced preferences), whereas children cannot. This sort of gene–environment correlation does not fit comfortably on either side of the genetic–environmental dichotomy! I cannot pursue these issues here; for more on behavior genetics, see Neisser et al. (1996) and the references cited there.

BLACK–WHITE DIFFERENCES

The Flynn gains have implications for a related problem, one that the authors of *The Bell Curve* regard as intractable. The 15-point "gap" between the IQ means of Blacks and Whites in the United States has persisted for many decades. This gap has serious social consequences:

The abilities measured by intelligence tests are important in the workplace as well as the school. Nevertheless, its cause still is not known. Some of the more obvious hypotheses have long been refuted: It does not result simply from racial bias in the language or cultural content of the tests. (In fact, the Black–White difference is somewhat larger on nonverbal tests.) Some currently plausible explanations refer to differences of caste and culture; there are also hypotheses based on early experience, nutrition, and the like. These hypotheses may seem plausible, but in this case, too, none of them are firmly established (for a review, see Neisser et al., 1996).

Some theorists have suggested that the persistent Black–White difference may have a genetic basis. Although the authors of *The Bell Curve* regard this hypothesis as plausible, there is little direct evidence to support it (Neisser, 1997a). In particular, the existence of genetic differences *within* both the White and the Black populations implies nothing one way or another for the difference between those populations. A well-known example given by Lewonton (1970) makes this point clear. Imagine that two fields of corn have been planted with the same strain of genetically varied seeds but that only one field is adequately watered and fertilized. The result will be an entirely environmental between-field difference, together with a large and entirely genetic within-field variance.

Whatever the merits of these various explanations, they may all soon be out of date. This is for two reasons. The first reason, of course, is the rise in test scores. As we have seen, the 3-point-per-decade gain documented by Flynn means that the test performance of Black Americans today is roughly equivalent to that of Whites in the 1940s. Even if the mean test scores are still 15 points apart, it is now clear that a gap of this size can easily result from environmental differences, specifically, from the differences between the general American environments of 1940 and 1990.

The second reason is that the gap itself may be closing. Where IQ scores themselves are concerned, the situation is not clear. Although some researchers have reported a recent reduction in Black–White IQ differences among children (Vincent, 1991), the samples have been small and

there are also data to the contrary (Lynn, 1996). Meanwhile, a more impressive convergence has appeared in a related domain: measures of children's school learning. This trend is the subject of Part 2 of this volume.

SCHOOL ACHIEVEMENT

IQ scores predict school achievement fairly well; that is what they were designed to do. Nevertheless, what children actually learn depends on many factors other than their intelligence. Some of those factors are other characteristics of the child, such as effort, attitude, and the like, but many are characteristics of the school environment. Other things being equal, how much mathematics a child learns will depend on how much time his or her school devotes to instruction in that subject, how much the importance of that instruction is emphasized, and so on. These emphases can vary from time to time and classroom to classroom as well as from country to country and school to school.

The most widely known index of school achievement is probably the SAT (recently renamed the Scholastic Assessment Test), but this test has many disadvantages from a scientific point of view. The most important of these is that it does not involve a representative sample: Children do not take the SAT unless they expect to go to college. A much better measure is the National Assessment of Educational Progress (NAEP), which is regularly given to appropriate national samples of children. NAEP data show that the average reading and math achievement of African American children went up substantially in the 1970s and 1980s, while those of White children did not. The Black–White gap in the math scores of 17-year-olds (e.g.) was about 1.1 standard deviation units as recently as 1978; by 1990, it had shrunk to about 0.6 units. Verbal scores showed similar trends. These gains were not limited to African Americans: Hispanic children showed comparable though somewhat smaller increases.

Most of the discussion in Part 2 centers on the NAEP data. (In chapter 12, Huang & Hauser report similar gains on a vocabulary scale that is regularly administered as part of a social survey.) The focus is

not simply on the existence of the gains, which is well established, but on their possible causes. What factors might have produced such substantial increases in so short a time? In chapter 10, David Grissmer et al. show that the gains partly reflect demographic shifts. Today's African American children tend to come from smaller families and to have better educated parents than was the case in earlier times; both of these factors are known to be associated with higher school achievement. Nevertheless, demographic shifts are not enough to account for all of the observed gains. Something else must be going on too, and these authors suggest a plausible candidate. During the 1960s and 1970s, the federal government adopted a wide range of policies designed to improve the education of Black children; these policies included, but were not limited to, school desegregation. Despite what is often assumed, those efforts seem to have been somewhat successful. The relative rise in Black children's school achievement is thus less mysterious than the general rise in IQ scores; we know at least some of its causes. Whether these phenomena are linked in any way remains to be seen.

THE HYPOTHESIS OF DYSGENIC TRENDS

Parts 1 and 2 of this book begin with established effects—the worldwide rise in IQ scores and the narrowing Black–White gap in school achievement—and consider their possible causes. That is, they start with results, and then speculate about "mechanisms." Part 3, in contrast, begins with a mechanism and goes on to speculate about its effects. The "mechanism" in question is the trend mentioned earlier: People of high intelligence (or at least of high IQ) tend to have fewer children than those of lower intelligence. Even though we are presently going through an environmentally driven rise in IQ scores, it remains possible that a genetic trend in the opposite direction is also taking place. This hypothesis is defended by Richard Lynn in chapter 13 and critically evaluated in chapters 14–16.

There are actually two steps in the argument. The first is to show that a "negative fertility differential" presently exists, that is, that people of low IQ have more surviving children than people of high IQ. Although Lynn presents a range of data to establish this point, such differentials can change rather quickly from one cultural era to the next. The second step is to prove that such a differential inevitably produces a decline in the average IQ of the whole population. Lynn, like Galton and others before him, simply takes this for granted. Samuel Preston shows in chapter 15, however, that it is a mistake to do so: Under many conditions, differential fertility rates are fully compatible with a stable population mean! Although the contributors to Part 3 do not reach a consensus on this and other issues, one can at least conclude that no dysgenic trend for intelligence has been conclusively demonstrated. If such a trend does exist, it is probably too small to be cause for alarm.

The Flynn effect, in contrast, is large indeed. It is also profoundly significant: The existence of 3-point-per-decade gains in IQ (and of even larger gains on *g*-loaded tests) will surely transform the intelligence debate. The authors of *The Bell Curve* assumed that the present (1990s) test scores of certain population groups set inevitable limits on what they and their children can attain. This assumption turns out to be false. The intellectual abilities of population groups are not carved forever in the genes; they can and do go up from one generation to the next. IQ, *g*, and school achievement are all massively affected by environmental factors. The next task for research and analysis is to understand what those factors are and how they have their effects. That understanding will set the stage for a new and more constructive debate, a discussion focused on the nature and meaning of the rising curve.

REFERENCES

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices test. *Psychological Review, 97,* 404–431.

Cronbach, L. J. (1970). *Essentials of psychological testing* (3rd ed). New York: Harper & Row.

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

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

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

Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.

Lewonton, R. (1970). Race and intelligence. *Bulletin of the Atomic Scientists, 26*, 2–8.

Lippmann, W. (1976). The mental age of Americans. In N. J. Block & G. Dworkin (Eds.), *The IQ controversy* (pp. 4–8). New York: Pantheon Books (Random House).

Lynn, R. (1996). Racial and ethnic differences in intelligence in the United States on the Differential Ability Scale. *Personality and Individual Differences, 20*, 271–273.

Neisser, U. (1997a). Never a dull moment. *American Psychologist, 52*, 79–81.

Neisser, U. (1997b). Rising scores on intelligence tests. *American Scientist, 85*, 440–447.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51*, 77–101.

Schaie, K. W. (1983). The Seattle Longitudinal Study: A twenty-one year exploration of psychometric intelligence in adulthood. In K. W. Schaie (Ed.), *Longitudinal studies of adult psychological development* (pp. 64–135). New York: Guilford Press.

Schaie, K. W. (1997). The course of adult intellectual development. *American Psychologist, 49*, 304–313.

Spearman, C. (1904). "General Intelligence" objectively determined and measured. *American Journal of Psychology, 15*, 201–293.

Tuddenham, R. (1948). Soldier intelligence in World Wars I and II. *American Psychologist, 3*, 54–56.

Vincent, K. R. (1991). Black/White IQ differences: Does age make a difference? *Journal of Clinical Psychology, 47*, 266–270.

Wechsler, D. (1981). *WAIS-R manual*. New York: Psychological Corporation.

Zajonc, R. B. (1976). Family configuration and intelligence. *Science, 192*, 227–236.

Zajonc, R. B., & Mullally, P. R. (1997). Birth order: Reconciling conflicting effects. *American Psychologist, 52*, 685–699.

PART ONE

Gains on Intelligence Tests