

from The New Yorker

December 17, 2007
BOOKS

None of the Above

What I.Q. Doesn't tell you about race.

by Malcolm Gladwell

If what I.Q. tests measure is immutable and innate, what explains the Flynn effect—the steady rise in scores across generations?

1.

One Saturday in November of 1984, James Flynn, a social scientist at the University of Otago, in New Zealand, received a large package in the mail. It was from a colleague in Utrecht, and it contained the results of I.Q. tests given to two generations of Dutch eighteen-year-olds. When Flynn looked through the data, he found something puzzling. The Dutch eighteen-year-olds from the nineteen-eighties scored better than those who took the same tests in the nineteen-fifties—and not just slightly better, much better.

Curious, Flynn sent out some letters. He collected intelligence-test results from Europe, from North America, from Asia, and from the developing world, until he had data for almost thirty countries. In every case, the story was pretty much the same. I.Q.s around the world appeared to be rising by 0.3 points per year, or three points per decade, for as far back as the tests had been administered. For some reason, human beings seemed to be getting smarter.

Flynn has been writing about the implications of his findings—now known as the Flynn effect—for almost twenty-five years. His books consist of a series of plainly stated statistical observations, in support of deceptively modest conclusions, and the evidence in support of his original observation is now so overwhelming that the Flynn effect has moved from theory to fact. What remains uncertain is how to make

sense of the Flynn effect. If an American born in the nineteen-thirties has an I.Q. of 100, the Flynn effect says that his children will have I.Q.s of 108, and his grandchildren I.Q.s of close to 120—more than a standard deviation higher. If we work in the opposite direction, the typical teenager of today, with an I.Q. of 100, would have had grandparents with average I.Q.s of 82—seemingly below the threshold necessary to graduate from high school. And, if we go back even farther, the Flynn effect puts the average I.Q.s of the schoolchildren of 1900 at around 70, which is to suggest, bizarrely, that a century ago the United States was populated largely by people who today would be considered mentally retarded.

2.

For almost as long as there have been I.Q. tests, there have been I.Q. fundamentalists. H. H. Goddard, in the early years of the past century, established the idea that intelligence could be measured along a single, linear scale. One of his particular contributions was to coin the word "moron." "The people who are doing the drudgery are, as a rule, in their proper places," he wrote. Goddard was followed by Lewis Terman, in the nineteen-twenties, who rounded up the California children with the highest I.Q.s, and confidently predicted that they would sit at the top of every profession. In 1969, the psychometrician Arthur Jensen argued that programs like Head Start, which tried to boost the academic performance of minority children, were doomed to failure, because I.Q. was so heavily genetic; and in 1994 Richard Herrnstein and Charles Murray, in "The Bell Curve," notoriously proposed that Americans with the lowest I.Q.s be sequestered in a "high-tech" version of an Indian reservation, "while the rest of America tries to go about its business." To the I.Q. fundamentalist, two

things are beyond dispute: first, that I.Q. tests measure some hard and identifiable trait that predicts the quality of our thinking; and, second, that this trait is stable—that is, it is determined by our genes and largely impervious to environmental influences.

This is what James Watson, the co-discoverer of DNA, meant when he told an English newspaper recently that he was "inherently gloomy" about the prospects for Africa. From the perspective of an I.Q. fundamentalist, the fact that Africans score lower than Europeans on I.Q. tests suggests an ineradicable cognitive disability. In the controversy that followed, Watson was defended by the journalist William Saletan, in a three-part series for the online magazine Slate. Drawing heavily on the work of J. Philippe Rushton—a psychologist who specializes in comparing the circumference of what he calls the Negroid brain with the length of the Negroid penis—Saletan took the fundamentalist position to its logical conclusion. To erase the difference between blacks and whites, Saletan wrote, would probably require vigorous interbreeding between the races, or some kind of corrective genetic engineering aimed at

upgrading African stock. "Economic and cultural theories have failed to explain most of the pattern," Saletan declared, claiming to have been "soaking [his] head in each 's computations and arguments." One argument that Saletan never soaked his head in, however, was Flynn's, because what Flynn discovered in his mailbox upsets the certainties upon which I.Q. fundamentalism rests. If whatever the thing is that I.Q. tests measure can jump so much in a generation, it can't be all that immutable and it doesn't look all that innate.

The very fact that average I.Q.s shift over time ought to create a "crisis of confidence," Flynn writes in "What Is Intelligence?" (Cambridge; \$22), his latest attempt to puzzle through the implications of his discovery. "How could such huge gains be intelligence gains? Either the children of today were far brighter than their parents or, at least in some circumstances, I.Q. tests were not good measures of intelligence."

3.

The best way to understand why I.Q.s rise, Flynn argues, is to look at one of the most widely used I.Q. tests, the so-called WISC (for Wechsler Intelligence Scale for Children). The WISC is composed of ten subtests, each of which measures a different aspect of I.Q. Flynn points out that scores in some of the categories—those measuring general knowledge, say, or vocabulary or the ability to do basic arithmetic—have risen only modestly over time. The big gains on the WISC are largely in the category known as "similarities," where you get questions such as "In what way are xdogs' and xrabbits' alike?" Today, we tend to give what, for the purposes of I.Q. tests, is the right answer: dogs and rabbits are both mammals. A nineteenth-century American would have said that "you use dogs to hunt rabbits."

"If the everyday world is your cognitive home, it is not natural to detach abstractions and logic and the hypothetical from their concrete referents," Flynn writes. Our great-grandparents may have been perfectly intelligent. But

they would have done poorly on I.Q. tests because they did not participate in the twentieth century's great cognitive revolution, in which we learned to sort experience according to a new set of abstract categories. In Flynn's phrase, we have now had to put on "scientific spectacles," which enable us to make sense of the WISC questions about similarities. To say that Dutch I.Q. scores rose substantially between 1952 and 1982 was another way of saying that the Netherlands in 1982 was, in at least certain respects, much more cognitively demanding than the Netherlands in 1952. An I.Q., in other words, measures not so much how smart we are as how modern we are.

This is a critical distinction. When the children of Southern Italian immigrants were given I.Q. tests in the early part of the past century, for example, they recorded median scores in the high seventies and low eighties, a full standard deviation below their American and Western European counterparts. Southern Italians did as poorly on I.Q. tests as Hispanics and blacks did. As you can imagine, there was much concerned talk at the time about the genetic inferiority of Italian stock, of the inadvisability of letting so many second-class

immigrants into the United States, and of the squalor that seemed endemic to Italian urban neighborhoods. Sound familiar? These days, when talk turns to the supposed genetic differences in the intelligence of certain races, Southern Italians have disappeared from the discussion. "Did their genes begin to mutate somewhere in the 1930s?" the psychologists Seymour Sarason and John Doris ask, in their account of the Italian experience. "Or is it possible that somewhere in the 1920s, if not earlier, the sociocultural history of Italo-Americans took a turn from the blacks and the Spanish Americans which permitted their assimilation into the general undifferentiated mass of Americans?"

The psychologist Michael Cole and some colleagues once gave members of the Kpelle tribe, in Liberia, a version of the WISC similarities test: they took a basket of food, tools, containers, and clothing and asked the tribesmen to sort them into appropriate categories. To the frustration of the researchers, the Kpelle chose functional pairings. They put a potato and a knife together because a knife is

used to cut a potato. "A wise man could only do such-and-such," they explained. Finally, the researchers asked, "How would a fool do it?" The tribesmen immediately re-sorted the items into the "right" categories. It can be argued that taxonomical categories are a developmental improvement—that is, that the Kpelle would be more likely to advance, technologically and scientifically, if they started to see the world that way. But to label them less intelligent than Westerners, on the basis of their performance on that test, is merely to state that they have different cognitive preferences and habits. And if I.Q. varies with habits of mind, which can be adopted or discarded in a generation, what, exactly, is all the fuss about?

When I was growing up, my family would sometimes play Twenty Questions on long car trips. My father was one of those people who insist that the standard categories of animal, vegetable, and mineral be supplemented with a fourth category: "abstract." Abstract could mean something like "whatever it was that was going through my mind when we drove past the water tower fifty miles back." That abstract

category sounds absurdly difficult, but it wasn't: it merely required that we ask a slightly different set of questions and grasp a slightly different set of conventions, and, after two or three rounds of practice, guessing the contents of someone's mind fifty miles ago becomes as easy as guessing Winston Churchill. (There is one exception. That was the trip on which my old roommate Tom Connell chose, as an abstraction, "the Unknown Soldier"—which allowed him legitimately and gleefully to answer "I have no idea" to almost every question. There were four of us playing. We gave up after an hour.) Flynn would say that my father was teaching his three sons how to put on scientific spectacles, and that extra practice probably bumped up all of our I.Q.s a few notches. But let's be clear about what this means. There's a world of difference between an I.Q. advantage that's genetic and one that depends on extended car time with Graham Gladwell.

4.

Flynn is a cautious and careful writer. Unlike many others in the I.Q. debates, he resists grand philosophizing. He comes back again and again to the fact that I.Q. scores are generated by paper-and-

pencil tests—and making sense of those scores, he tells us, is a messy and complicated business that requires something closer to the skills of an accountant than to those of a philosopher.

For instance, Flynn shows what happens when we recognize that I.Q. is not a freestanding number but a value attached to a specific time and a specific test. When an I.Q. test is created, he reminds us, it is calibrated or "normed" so that the test-takers in the fiftieth percentile—those exactly at the median—are assigned a score of 100. But since I.Q.s are always rising, the only way to keep that hundred-point benchmark is periodically to make the tests more difficult—to "renorm" them. The original WISC was normed in the late nineteen-forties. It was then renormed in the early nineteen-seventies, as the WISC-R; renormed a third time in the late eighties, as the WISC III; and renormed again a few years ago, as the WISC IV—with each version just a little harder than its predecessor. The notion that anyone "has" an I.Q. of a certain number, then, is meaningless unless you know which WISC he took, and when he took it, since there's a substantial difference between getting a

130 on the WISC IV and getting a 130 on the much easier WISC.

This is not a trivial issue. I.Q. tests are used to diagnose people as mentally retarded, with a score of 70 generally taken to be the cutoff. You can imagine how the Flynn effect plays havoc with that system. In the nineteen-seventies and eighties, most states used the WISC-R to make their mental-retardation diagnoses. But since kids—even kids with disabilities—score a little higher every year, the number of children whose scores fell below 70 declined steadily through the end of the eighties. Then, in 1991, the WISC III was introduced, and suddenly the percentage of kids labelled retarded went up. The psychologists Tomoe Kanaya, Matthew Scullin, and Stephen Ceci estimated that, if every state had switched to the WISC III right away, the number of Americans labelled mentally retarded should have doubled.

That is an extraordinary number. The diagnosis of mental disability is one of the most stigmatizing of all educational and occupational classifications—and yet, apparently, the chances of being burdened

with that label are in no small degree a function of the point, in the life cycle of the WISC, at which a child happens to sit for his evaluation. "As far as I can determine, no clinical or school psychologists using the WISC over the relevant 25 years noticed that its criterion of mental retardation became more lenient over time," Flynn wrote, in a 2000 paper. "Yet no one drew the obvious moral about psychologists in the field: They simply were not making any systematic assessment of the I.Q. criterion for mental retardation."

Flynn brings a similar precision to the question of whether Asians have a genetic advantage in I.Q., a possibility that has led to great excitement among I.Q. fundamentalists in recent years. Data showing that the Japanese had higher I.Q.s than people of European descent, for example, prompted the British psychometrician and eugenicist Richard Lynn to concoct an elaborate evolutionary explanation involving the Himalayas, really cold weather, premodern hunting practices, brain size, and specialized vowel sounds. The fact that the I.Q.s of Chinese-Americans also seemed to be elevated has led I.Q. fundamentalists to posit the existence of an international

I.Q. pyramid, with Asians at the top, European whites next, and Hispanics and blacks at the bottom.

Here was a question tailor-made for James Flynn's accounting skills. He looked first at Lynn's data, and realized that the comparison was skewed. Lynn was comparing American I.Q. estimates based on a representative sample of schoolchildren with Japanese estimates based on an upper-income, heavily urban sample. Recalculated, the Japanese average came in not at 106.6 but at 99.2. Then Flynn turned his attention to the Chinese-American estimates. They turned out to be based on a 1975 study in San Francisco's Chinatown using something called the Lorge-Thorndike Intelligence Test. But the Lorge-Thorndike test was normed in the nineteen-fifties. For children in the nineteen-seventies, it would have been a piece of cake. When the Chinese-American scores were reassessed using up-to-date intelligence metrics, Flynn found, they came in at 97 verbal and 100 nonverbal. Chinese-Americans had slightly lower I.Q.s than white Americans.

The Asian-American success story had suddenly been

turned on its head. The numbers now suggested, Flynn said, that they had succeeded not because of their higher I.Q.s. but despite their lower I.Q.s. Asians were overachievers. In a nifty piece of statistical analysis, Flynn then worked out just how great that overachievement was. Among whites, virtually everyone who joins the ranks of the managerial, professional, and technical occupations has an I.Q. of 97 or above. Among Chinese-Americans, that threshold is 90. A Chinese-American with an I.Q. of 90, it would appear, does as much with it as a white American with an I.Q. of 97.

There should be no great mystery about Asian achievement. It has to do with hard work and dedication to higher education, and belonging to a culture that stresses professional success. But Flynn makes one more observation. The children of that first successful wave of Asian-Americans really did have I.Q.s that were higher than everyone else's—coming in somewhere around 103. Having worked their way into the upper reaches of the occupational scale, and taken note of how much the professions value abstract thinking, Asian-American parents have

evidently made sure that their own children wore scientific spectacles. "Chinese Americans are an ethnic group for whom high achievement preceded high I.Q. rather than the reverse," Flynn concludes, reminding us that in our discussions of the relationship between I.Q. and success we often confuse causes and effects. "It is not easy to view the history of their achievements without emotion," he writes. That is exactly right. To ascribe Asian success to some abstract number is to trivialize it.

5.

Two weeks ago, Flynn came to Manhattan to debate Charles Murray at a forum sponsored by the Manhattan Institute. Their subject was the black-white I.Q. gap in America. During the twenty-five years after the Second World War, that gap closed considerably. The I.Q.s of white Americans rose, as part of the general worldwide Flynn effect, but the I.Q.s of black Americans rose faster. Then, for about a period of twenty-five years, that trend stalled—and the question was why.

Murray showed a series of PowerPoint slides, each representing different statistical formulations of the I.Q. gap. He appeared to be

pessimistic that the racial difference would narrow in the future. "By the nineteen-seventies, you had gotten most of the juice out of the environment that you were going to get," he said. That gap, he seemed to think, reflected some inherent difference between the races. "Starting in the nineteen-seventies, to put it very crudely, you had a higher proportion of black kids being born to really dumb mothers," he said. When the debate's moderator, Jane Waldfogel, informed him that the most recent data showed that the race gap had begun to close again, Murray seemed unimpressed, as if the possibility that blacks could ever make further progress was inconceivable.

Flynn took a different approach. The black-white gap, he pointed out, differs dramatically by age. He noted that the tests we have for measuring the cognitive functioning of infants, though admittedly crude, show the races to be almost the same. By age four, the average black I.Q. is 95.4—only four and a half points behind the average white I.Q. Then the real gap emerges: from age four through twenty-four, blacks lose six-tenths of a point a year, until their scores settle at 83.4.

That steady decline, Flynn said, did not resemble the usual pattern of genetic influence. Instead, it was exactly what you would expect, given the disparate cognitive environments that whites and blacks encounter as they grow older. Black children are more likely to be raised in single-parent homes than are white children—and single-parent homes are less cognitively complex than two-parent homes. The average I.Q. of first-grade students in schools that blacks attend is 95, which means that "kids who want to be above average don't have to aim as high." There were possibly adverse differences between black teen-age culture and white teen-age culture, and an enormous number of young black men are in jail—which is hardly the kind of environment in which someone would learn to put on scientific spectacles.

Flynn then talked about what we've learned from studies of adoption and mixed-race children—and that evidence didn't fit a genetic model, either. If I.Q. is innate, it shouldn't make a difference whether it's a mixed-race child's mother or father who is black. But it does: children with a white mother and a black father have an eight-point I.Q. advantage over those with a

black mother and a white father. And it shouldn't make much of a difference where a mixed-race child is born. But, again, it does: the children fathered by black American G.I.s in postwar Germany and brought up by their German mothers have the same I.Q.s as the children of white American G.I.s and German mothers. The difference, in that case, was not the fact of the children's blackness, as a fundamentalist would say. It was the fact of their Germanness—of their being brought up in a different culture, under different circumstances. "The mind is much more like a muscle than we've ever realized," Flynn said. "It needs to get cognitive exercise. It's not some piece of clay on which you put an indelible mark." The lesson to be drawn from black and white differences was the same as the lesson from the Netherlands years ago: I.Q. measures not just the quality of a person's mind but the quality of the world that person lives in.

© 2007 Malcolm Gladwell