# 1 Opening windows

Are we getting smarter? If you mean "Do our brains have more potential at conception than those of our ancestors?" then we are not. If you mean "Are we developing mental abilities that allow us to better deal with the complexity of the modern world, including problems of economic development?" then we are. For most people, the latter is what counts, so I will let the affirmative answer stand. But scholars prefer to ask a different question, to which they attach a special meaning, namely "Are we getting more intelligent?" I will answer that question at the end of Chapter 2.

Whatever we are doing, we are making massive IQ gains from one generation to another. That in itself is of great significance. IQ trends over time open windows on the human condition that make us conscious of things of which we were only half aware. This book attempts to make sense of what time and place are doing to our minds. It has new things to say about cognitive trends in both the developed and the developing world and where they may go over the rest of this century. It falsifies a major hypothesis that suggests that IQ differences between the two worlds are set in the stone of genetic differences. It addresses the most recent debate about the death penalty, particularly attempts to obscure the relevance of IQ gains to who lives or dies. It shows that cognitive trends have rendered inaccurate the diagnosis of memory loss. Perhaps most disturbing, it adds a new dimension to the tendency of western adults and teenagers to grow apart since 1950.

However, all the topics covered do not fit neatly into the box of IQ trends over time. I have included new thinking and data on subjects of general interest: whether race and gender IQ differences are genetic or environmental in origin; how modernity (or lack of it) affects the cognitive abilities of women; whether old age has a darker side hitherto unperceived. And finally, I offer a diagnosis suggested by some 30 years in the field: that psychology has somehow drifted away from sociology and suffered thereby.

Five years ago I published *What Is Intelligence? Beyond the Flynn Effect* (2007) and updated it two years later in the expanded paperback edition (2009). I thought of updating it again. However, as indicated, my new thinking and discoveries did not advance the theory of intelligence so much as a whole range of issues concerning economic growth, the death penalty, aging, and group differences.

Nonetheless what was said in the previous book colors my approach and therefore, the next chapter summarizes its contents. I do not flatter myself that everyone who reads this will have read (or will want to read) my previous work. Still, even those who have read *What Is Intelligence?* may find the next chapter interesting. It gives, for the first time, a full account of adult gains on the WAIS (Wechsler Adult Intelligence Scale), and compares them to child gains on the WISC (Wechsler Intelligence Scale for Children). Moreover, when a book is condensed, connections emerge that were not so clear in the lengthy original.

As for the remainder of this book, Chapter 3 speculates about the distant past and cognitive trends over the rest of this century. It also traces trends on Raven's Progressive Matrices in the UK over the last 65 years, and gives a final verdict on the role of nutrition. Chapter 4 criticizes those who make *Daubert* motions, so they can go on using uncorrected IQ scores to multiply death sentences. It also presents evidence that instruments

in current use misdiagnose memory loss in both Britain and Sweden.

Chapter 5 looks at American vocabulary trends over the last half-century. It assesses whether adult gains are the result of the spread of tertiary education or the expansion of cognitively demanding work, and notes a worrying trend for the language used by parents and the language used by their children to diverge. It also analyzes whether high-IQ or low-IQ people are more at risk of a radical loss of cognitive ability in old age. The evidence suggests that while there is a bonus for being bright in retaining vocabulary, there is a levy on being bright in retaining analytic skills.

Chapter 6 argues that the differential performance of black and white Americans on Wechsler subtests does not indicate whether the black/white IQ gap is genetic or environmental in origin. It also shows that modern women match men on Raven's Progressive Matrices, despite the fact that university women have a slightly lower IQ than university men.

Chapter 7 argues that something peculiar happens to the study of intelligence when it becomes sociologically blind. Chapter 8, the last chapter, offers a brief summary and ends with a tribute to *g* and Arthur Jensen.

# 2 IQ and intelligence

Whether the twentieth century has seen intelligence gains is controversial. Whether there have been massive IQ gains over time is not. I will: (1) describe the range and pattern of IQ gains; (2) discuss their historical and social significance; (3) argue that they suggest a new theory of intelligence; and (4) urge that understanding them is more important than classifying them (as either intelligence or non-intelligence gains).

## The evidence and its peculiarities

Reed Tuddenham (1948) was the first to present convincing evidence of massive gains on mental tests using a nationwide sample. He showed that US soldiers had made about a 14-point gain on Armed Forces tests between World War I and World War II or almost a full standard deviation (SD = 15 throughout). The tests in question had a high loading on the kind of material taught in the classroom, and he thought the gains were primarily a measure of improved schooling. Therefore, they seemed to have no theoretical implications, and because the tests were not among those most often used by clinical psychologists the practical implications were ignored.

When Flynn (1984, 1987) showed that massive gains had occurred in America on Wechsler and Stanford–Binet IQ tests, and that they had occurred throughout the industrialized world even on tests thought to be the purest measures of intelligence, IQ gains took center stage. Within a decade, Herrnstein and

Murray (1994), the authors of *The Bell Curve*, called the phenomenon the "Flynn effect."

Nations with data about IQ trends stand at 31. Scandinavian nations had robust gains but these peaked about 1990 and since then, may have gone into mild decline. Several other nations show persistent gains. Americans are still gaining at their historic rate of 0.30 points per year (WAIS 1995–2006; WISC 1989–2002). British children were a bit below that on Raven's from 1980 to 2008, but their current rate of gain is higher than in the earlier period from 1943 to 1980. German adults were still making vocabulary gains in 2007 at a slightly higher rate than US adults. South Korean children gained at double the US rate between 1989 and 2002 (Emanuelsson, Reuterberg, & Svensson, 1993; Flynn, 2009a, 2009b; Pietschnig, Voracek, & Formann, 2010; Schneider, 2006; Sundet, Barlaug, & Torjussen, 2004; Teasdale & Owen, 1989, 2000; te Nijenhuis, 2011; te Nijenhuis *et al.*, 2008).

Other recent gains cover long periods, so whether the rate varied approaching the present is unknown. Urban Argentines (ages 13 to 24) made a 22-point gain on Raven's between 1964 and 1998. Children in urban Brazil (1930–2002), Estonia (1935–1998), and Spain (1970–1999) made gains akin to the US rate. All in all, gains from the developed world cover the United States; 15 European nations or peoples; four Asian nations (urban China, India, Japan, and South Korea); three Commonwealth nations (Australia, Canada, and New Zealand); urban Brazil and urban Argentina; Israel; and white South Africa (Colom, Flores-Mendoza, & Abad, 2007; Colom, Lluis Font, & Andres-Pueyo, 2005; Flynn, 1987, 1998b, 2009c; Flynn & Rossi-Casé, 2011 ; Murphy, te Nijenhuis, & van Eeden, 2008; Must, Must, & Raudik, 2003; te Nijenhuis, 2011).

The developing world has begun to show explosive gains in rural Kenya, Dominica, Turkey, and Saudi Arabia. In Sudan, large gains on the WAIS Performance Scale were accompanied

by a small loss for tests closer to school learning (Batterjee *et al.*, in press; Daley *et al.*, 2003; Kagitcibasi, & Biricik, 2011; Khaleefa, Afra Sulman, & Lynn, 2009; Meisenberg *et al.*, 2005).

The Dutch data made the greatest impact. Between 1952 and 1982, young Dutch males gained 20 IQ points on a test of 40 items selected from Raven's Progressive Matrices (Flynn, 1987). The sample was exhaustive. Raven's was supposed to be the exemplar of a culturally reduced test, one that should have shown no gains over time as culture evolved. These 18-year-olds had reached the age at which performance on Raven's peaks. Therefore, their gains could not be dismissed as early maturation, that is, it was not just a matter that children today matured a few years earlier than the children of yesterday. Current people would have a much higher IQ than the last generation even after both had reached maturity.

The Dutch gains created a crisis of confidence. How could such huge gains be intelligence gains? The gains amounted to 1.33 SDs. This would put the average Dutch person of 1982 at the 90th percentile of Dutch in 1952. Psychologists faced a paradox: either the people of today were far brighter than their parents or, at least in some circumstances, IQ tests were not good measures of intelligence.

Box 1 shows how large American gains have been on the most frequently used tests, namely, the WISC (Wechsler Intelligence Scale for Children) and the WAIS (Wechsler Adult Intelligence Scale). These show Full Scale IQ gains proceeding at 0.30 points per year over the last half of the twentieth century, a rate often found in other nations, for a total gain of over 15 points. If we link these to earlier data, such as that of Tuddenham, the gain over the last century has been at least 30 points.

The Dutch gains on Raven's run at over 0.60 points per year, double the rate for Wechsler tests. This is the case for most nations, at least at the time of their peak gains, and focuses us on how IQ tests differ. Raven's measures what is

**Box 1**

The magnitude of US gains on Wechsler tests for both children (WISC) and adults (WAIS) have been comparable, at least for Full Scale IQ. Setting IQs at 100 at the beginning of the period the data cover:

| | | | |
|---|---|---|---|
| WISC: 100.00 (1947–48) | 107.63 (1972) | 113.00 (1989) | 117.63 (2002) |
| WAIS: 100.00 (1953–54) | 107.50 (1978) | 111.70 (1995) | 115.07 (2006) |

*Sources:* Flynn, 2009b, 2009c, 2010.

called *fluid intelligence*, solving problems on the spot. You have to identify the missing piece of a design based on its logic, rather like (although often more demanding than) identifying the missing number in a series, say 2-4-8-10 (6 is missing). The Wechsler tests measure *crystallized intelligence*, which is knowledge of a sort you could not acquire unless you were capable of absorbing certain concepts; for example, you could not attain a large vocabulary unless you were good at grasping the concepts behind words. International Raven's data suggest that people have gained 50 points over the twentieth century. It has one rival. The Wechsler test battery consists of 10 subtests, ranging from vocabulary to three-dimensional jigsaw puzzles. One subtest shows gains near the magnitude of Raven's gains. It is the similarities subtest, which tests your ability to classify things that have something in common (e.g. dogs and rabbits are both mammals).

The pattern of IQ gains over time has a final peculiarity, namely, it is not factor-invariant (Wicherts *et al.*, 2004). Factor analysis is a technique that measures the extent to which those who excel on some IQ subtests also excel on others. The tendency toward general excellence is not peculiar to cognition. Just as those with larger vocabularies also tend to be better at

arithmetical reasoning and solving matrices problems, so people who are good at one musical instrument are often good at another, and people good at one sport are often good at almost all sports. When a variety of cognitive skills tend to intercorrelate, the measure of the tendency is called *g* (the general intelligence factor).

If the rank order of people on all subtests of the WISC were identical (one person topped them all, another person was second on them all, etc.), *g* would "explain" most of the pattern of test performance and have a high value, perhaps 0.80. If a person's score on each subtest were no more of an indication of their performance on any other subtest than a score chosen at random, *g* would be low or perhaps nil.

One subtest may have a higher "*g*-loading" than another. This means that it is a better guide as to who will do well on the other subtests. For example, if you added an eleventh WISC subtest on shoe tying, it would have a *g*-loading of close to zero: how fast you tie your shoes would have little relation to the size of your vocabulary. On the other hand, your score on the vocabulary subtest might be a pretty good predictor of your scores on the other subtests (except shoe tying) and get a *g*-loading of 0.75. You can rank the subtests into a hierarchy according to the size of their *g*-loadings.

## A pause to make a point

When tests or subtests are ranked according to their *g*-loadings, the skills with the greatest cognitive complexity tend to top the hierarchy, which is to say that the more complex the task, the more high-IQ people open up a gap on the average person. This is an intuitive judgment in that we have only our sense of what is complex to rely upon. But there are enough clear cases to establish the connection.

Imagine I was trying to convince someone that the intensity of heat was correlated with thermometer readings (and lacked a sophisticated knowledge of the science, which I do). I would first choose clear cases; for example I would choose pairs of days during which the temperature had obviously risen and say, "You see that the thermometer shows that it is 10 degrees (Fahrenheit) hotter than it was yesterday." After several such demonstrations, I would urge him to trust the thermometer on days that were close calls, days on which we disagreed about whether or not it was a bit hotter than it was yesterday. Sometimes he would be right, of course, which would fortify his confidence.

There are many clear cases in which differences of cognitive complexity are caught by differences in *g*-loadings. Making a soufflé is more *g*-loaded than scrambling eggs. Digit span backward (repeating numbers in the reverse order you heard them) is more *g*-loaded than digit span forward (repeating numbers in the same order you heard them). Coding (simply pairing symbols and numbers) has by far the lowest *g*-loading of all the Wechsler subtests. Mental arithmetic is far more *g*-loaded than when you are allowed to do the mathematics with a calculator. When we coach people to take IQ tests, we reduce problems that make them think on their feet to problems they can solve merely by applying a method they have been taught; and the *g*-loading falls dramatically.

Its correlation with cognitive complexity gives *g* a good case to be identified with intelligence. If you are still unconvinced, imagine that there were lower *g*-loadings for making soufflés and digit span backward and so forth. Surely this would falsify the claim of *g* to represent intelligence (or at least a certain kind of intelligence). Jensen goes on to suggest that there might exist a latent trait, general intelligence; and that to the extent to which a person possesses that trait the better he or she will do on a whole range of cognitive tasks.

We can now understand why it is thought significant that IQ gains are not consistently factor invariant. As far as *g* is concerned, this means that when we rank subtests by their *g*-loadings, we find that the magnitude of IQ gains on the various subtests do not tally. The largest IQ gain over time may be on a subtest with an average *g*-loading, with a smaller gain on the subtest with the highest *g*-loading. This convinced Jensen (1998) that the bulk of IQ gains were not *g* gains and therefore, were not intelligence gains. He suggests that IQ gains may be largely "hollow"; that is, they are a bundle of subtest-specific skills that have little real-world significance.

## Two kinds of significance

Before we accept the interpretation of IQ gains as hollow, it is useful to supplement factor analysis with functional analysis. Factor analysis may disclose latent traits but no one can do latent traits. What we do in the real world is perform, better or worse, functional activities, such as speaking, solving arithmetic problems, and reasoning about scientific and moral questions. To contrast the two, I will use a sports analogy.

If we factor analyze performances on the 10 events of the decathlon, a general factor or *g* would emerge and very likely subordinate factors representing speed (the sprints), spring (jumping events), and strength (throwing events). We would get a *g* because at a given time and place, performance on the 10 events would be intercorrelated; that is, someone who tended to be superior on any one would tend to be above average on all. We would also get various *g*-loadings for the 10 events, that is, superior performers would tend to rise further above average on some of them than on the others. The 100 meters would have a much higher *g*-loading than the 1,500 meters, which involves an endurance factor not very necessary in the other events.