Intelligence 41 (2013) 851–857

Contents lists available at ScienceDirect

Intelligence

The "Flynn Effect" and Flynn's paradox

James R. Flynn

University of Otago

1. Lynn on naming

Richard is correct. Calling massive IQ gains over time the "Flynn Effect" was an accident of history, a label Charles Murray coined in *The Bell Curve* in 1994. It is not a verdict a court would have been likely to hand down if it had an eye for the historical record. It is only fair to note that Murray's rationale did not attribute historical priority: "We call it 'the Flynn effect' because of psychologist James Flynn's pivotal role in focusing attention on it, but the phenomenon was identified in the 1930s when testers began to notice that IQ scores rose with every successive year after a test was first standardized." (Herrnstein & Murray, 1994, p. 307).

I have never administered an IQ test. Therefore, I was dependent on scholars who had administered them and recorded that their subjects scored higher IQs on older norms than on recent ones. What I could not understand in 1982, when I stumbled on the phenomenon, was why there was no fighting in the streets about it. Therefore, I wrote my 1984 article to show that the gains permeated the whole Wechsler–Binet literature, that the US gains were huge, and that they constituted an *unrecognized confounding variable* for literally hundreds of studies (Flynn, 1984). The response from Arthur Jensen was to tell me he had never realized what a mess the testers had made of things.

Jensen also challenged me as to whether the same phenomenon had occurred on culturally reduced tests like Raven's. Therefore, I published my 1987 article to show that it was a worldwide occurrence and that Raven's was affected more than any other test (Flynn, 1987). John Raven wrote that he had never realized how huge the gains were. Ian Deary told me that he was present at a lunch in 1987 where Jensen and others read the article with consternation. That article, of course, made explicit the dilemmas massive gains posed for the *theory of intelligence*. Were we getting that much smarter or were our ancestors mentally retarded?

My contribution, such as it is, was to bring the phenomenon to center stage; and subsequently, argue that it implied a new theory of intelligence that gave "*g*" an altered role. It was to be honored for its measurement of individual differences but had to make way for new concepts that were needed to write cognitive history (Flynn, 2009). Whether the latter was a contribution depends on whether people think my interpretation is correct. But the articles did focus the attention of psychologists and many others. Fortunately, however historically questionable, the label "Flynn effect" in 1994 engendered a whole new wave of interest. Perhaps even scholars need a phenomenon to be personalized to take notice? Therefore, I give my thanks to Charles Murray and my apologies to Richard Lynn.

I take issue, of course, with Lynn's interpretation of massive IQ gains (Flynn, 2009, 2012). The fact that gains begin at an early age is easily explained by cultural influences: hothouse parenting for small children and the advent of preschool. Thanks to starting school a year earlier, the contemporary child has two years of school compared to one at the age of 6, but only 6 years of school compared to 5 at the age of 10. This accounts for the fact that there is a tendency for the gains of the very young to diminish with age. But later they recover and even escalate. There are plenty of big gains for high school students and the huge gains for adults are a matter of record (which Lynn does not confront). My own theory identifies the causes that have made gains on non-verbal tests higher than those on verbal tests. As he notes, I have offered what I consider to be crushing evidence against nutrition as a sufficient cause and against its contributing much at all in advanced nations since 1950.

2. Dutton and Lynn

The introduction cites studies that show IQ gains have ended or gone into mild reverse in a number of nations. I consider Australia ambiguous until we get further data. Cotton et al. (2005) tested a normative sample on Raven's from Victoria only. The negative results conflict with Nettelbeck and Wilson (2004). The latter tested a (less representative) sample on the Peabody (PPVT) and it was from South Australia only. Woodley and Meisenberg (2013) show that the huge Dutch gains are over. I think it premature to generalize about the UK, given that Raven's gains in 2008 were large for all age groups except those aged 14.5–15.5 (Flynn, 2012, p.46). However, the new military data strongly

E-mail address: jim.flynn@otago.ac.nz.

0160-2896/$ – see front matter
http://dx.doi.org/10.1016/j.intell.2013.06.014

indicate that in 1997, Finland joined its Scandinavian neighbors in seeing off the IQ gains era.

The failure of massive gains to tail off throughout the developed world puzzles me. No gains would certainly help developing nations to catch the developed world (look ahead). The authors attribute differential patterns of IQ trends to low-IQ non-European immigration, excepting Finland. America is another exception: large immigration of lower-IQ Hispanics has not yet weakened its Wechsler rate of three IQ points per decade. Rindermann and Thompson (herein) find the US rate weak in the NAEP school data. I emphasize that it is no lower than gains on the WISC subtests that correspond to the school tests. Three German-speaking nations match US adult Vocabulary gains but only German speakers were included. South Korea is untouched by immigration and its gains are very large (Flynn, 2012, p. 37).

It is noteworthy that gains have ended in societies all of which are politically progressive. They may have achieved the full modernity that would weaken the triggers of IQ gains (no more progress to be made in education, cognitively demanding jobs and leisure, family size, visual saturation, and so forth).

3. Shiu, W., A. Beaujean, A., Must, O., te Nijenhuis, J., & Must, A.

Between 1934 and 2006, secondary school students in Estonia gained: 15.75 points on a synonyms–antonyms test; 11.85 points on vocabulary; an average of 12.4 points on several logic and analogies tests; 6.60 points on information; and essentially nil on arithmetic and computation. The large vocabulary gains are not atypical of continental Europe but far exceed those of American children. It would be wrong to interpret these gains as evidence that Estonia is still in the massive IQ gains phase because the period is so long and does not isolate trends over the last decade.

4. Jan te Nijenhuis and Henk van der Flier

The results appear to me correct: the magnitude of white/black IQ differences on Wechsler subtests at any given time is correlated with the *g* loadings of the subtests; the magnitude of IQ gains over time on subtests is not usually so correlated; the causes of the two phenomena are not the same. I have acknowledged this many times (Flynn, 2008, p. 79; 2012, p.136). The authors do not draw the following implications but for clarity's sake, I wish to make clear that there are certain implications that do not follow.

To think that the fact that black/white differences tally with *g* proves that the differences are genetic in origin is a mistake. Flynn (2008, pp. 88–92) analyzes the offspring of children fathered by either blacks or whites during the German occupation after World War II. Whatever subtest differences remained had no correlation with the *g* loadings of the subtests. I nominate the absence of a black subculture in Germany as a probable factor. The significance of this finding has nothing to do with whether there is a genetic component in the black/white IQ difference. There are many confounds that make the German data non-conclusive in this regard, for example, that the black fathers were a mild IQ elite. The point is that that while black genes may dictate a racial IQ gap, they do not dictate the *g* phenomenon. Flynn (2012, pp. 132–134) also uses a scenario to show that whenever one group lags behind another for cognitive complexity (*g*), environment can cause a *g* difference between them.

To think that the fact that the IQ gains of one race on another should tally with *g* (that the magnitude of race gains on a subtest should tally with the subtest *g* loadings), if they are to be significant, is a mistake. Its advocates must assert three propositions: the black/white IQ gap in 1972 was significant and real (thanks to tallying with *g*); the black/white IQ gap of 2002 was significant and real (thanks to tallying with *g*); and the IQ gains that reduced the gap were hollow and unreal (thanks to not tallying with *g*). It remains to be explained how a hollow trend could make a real-world difference (Flynn, 2013, p 26). Here I will insert a brief interjection.

5. Flynn's paradox

A question to which I hope every reader will respond: if the gains of one race on another need not tally with *g* to be significant, why must the gains of one generation on another do so? This time I will attach my name to a paradox about a prevalent use of *g*: unreal gains alter the real world (or) non-intelligence gains have intelligence effects (or) the necessity of *g* shows that *g* is unnecessary. My solution is simple: Stop using *g* as a kind of holy water that saves or damns.

6. Rindermann and Thompson

NAEP (National Assessment of Educational Progress) data from America cover 1971 to 2008 and offer highly reliable samples for ages 9, 13, and 17 on tests of reading and mathematics. The total reading gain (equivalent to 2 IQ points) is close to WISC Vocabulary gains over the same period. The mathematics gains for all ages average higher than WISC Arithmetic but reduces to only 1 point by age 17, which is comparable. I argue that the NAEP age pattern is a matter of mastering the mechanics of computation at earlier ages, but being no better at mathematical reasoning by age 17. The WISC Arithmetic subtest asks you to plan a strategy for solving problems and to do so using mental arithmetic. Thus I conclude that the negligible reasoning gains of the 17-year olds confirm the negligible strategy gains on the WISC (Flynn, 2009, pp. 21–22).

There has been a shift to gains over the lower half of the curve, partially due to the fact that low-IQ groups like blacks and Hispanics have made larger gains than whites. Between 1971 and 2008, averaging the scores for reading and mathematics, blacks gained 6.39 IQ points with the final gap for all ages at 9.94 points. Using all Wechsler and Stanford–Binet standardization samples, Dickens and Flynn (2006) concluded that between 1972 and 2002, blacks gained 5.5 IQ points on whites and that the average gap had fallen to 10.0 points for ages 9 to 17. The two data sets offer remarkably similar results. The Dickens and Flynn data cover all ages between 4 and 24 and sadly, black IQ steadily loses ground on white IQ as children age. In 2002 the gap was only 4.6 points at age 4 rising to 16.6 points at age 24.

7. Woodley, Figueredo, Brown, and Ross

I am gratified at the basic findings, which I interpret as follows: massive IQ gains over time have social significance

despite not correlating positively with the *g* loadings of various tests or subtests and despite having to overcome probable dysgenic trends; the Dickens–Flynn model can explain how "small" exogenous environmental factors can trigger large effects.

The social significance in this case is very interesting: whatever reinforces long-term strategies of executive function or problem solving is likely to be "beneficial". In other words, there is a correlation between *cognitive abilities* whose autonomous development allows one to cope with an increasingly complex modern world and *personality* indexes showing that the total human being has made a successful adaptation to the modern world. I have two reservations.

The failure of IQ trends to correlate with factor-analysis *g* does not mean that IQ gains are "independent of the population level of *g*" when we give *g* social significance. The notion of *g* would have no social significance if it did not correlate with cognitive complexity: if Raven's did not have a higher *g* loading than shoe-tying, or digit span backward did not have a higher loading than digit span forward. Since all the Wechsler subtests pose problems of cognitive complexity, even an "autonomous" set of gains on them (ones not mimicking the hierarchy of *g* loadings) means that people have advanced in terms of their ability to solve cognitively complex problems. Flynn (2013, pp. 27–29) borrows a method from the Woodcock–Johnson, which weights IQ gains in terms of *g*-loadings, and shows that "true" gains on *g* (problem solving ability) are almost identical to the magnitude of IQ gains.

The factors that measure what the authors call slow life-history speed (and that show positive results) include abstract thinking, creative writing and drawing, a personality-type (happiness, sociability, trust, self-control, enlightened self-interest), and longevity with greater control over fertility and ensuring infant survival. Since IQ gains signal cognitive adaptation to modernity, it makes sense that these correlates show that such a person has adapted successfully to modernity.

Historically no one might have thought to isolate these factors unless imbued with a life-history approach. But in theory, even a B. F. Skinner could ask whether cognitive "style" tallied with good life-history outcomes. He would have used his own language of course: what schedules of reinforcement throughout life encouraged long-term strategies and discouraged short-term ones. The Harvard sociologist Orlando Patterson (not a Skinnerian) has asked what cultural factors make black youth unlikely to pursue slow strategies (unlikely to study at school as a means of upward social mobility) and likely to pursue fast strategies of gaining prestige, such as sexual conquest, dressing "sharp", and hanging around in shopping malls. There is no doubt that the cultural milieu of black youth not only limits the diversified cognitive development needed to cope with modernity but also blights their prospects.

Despite these reservations, I should add that Woodley and his colleagues are doing some of the most important work on clarifying the effects of massive IQ gains.

8. Meisenberg and Woodley

This study is of outstanding quality. It is the first to systematically test the hypothesis that developing nations are likely to match the mean IQs of developed nations during the 21st century. Some of the former appear to be entering the "first phase" of modernity (massive gains) that the latter enjoyed last century. The predicted dates for a totally closed gap are: 40 years (about 2050) for PISA (math, science, and reading) and 341 years or perhaps never for TIMSS (math and science). The PISA data include few developing nations (mainly five from Latin America), while the TIMSS includes more of them particularly from Africa and the Middle East.

After studying six developing nations in some depth (Flynn, 2012, pp. 55–65), I predict that the 21st century will reflect both of these results. Some nations like Brazil, Turkey, and Kenya appear to have begun the road to modernity, while others like Sudan have not, and others like Saudi Arabia and Dominica have made some steps along the way but face daunting obstacles. My prediction is conditional on two things: continued progress uninterrupted by climate disaster and the trend toward nil IQ gains in developing nations becoming general. As to whether the developed versus developing gap will ever close, Lynn has speculated that there is a genetic hierarchy, one that entails that other nations will never rival the mean IQs of East Asians. This posits an evolutionary scenario (Chinese north of the Himalayas during the Ice Ages) that has been falsified (Flynn, 2012, pp. 33–35; Flynn, 2013, pp. 52–54).

It is particularly significant that rapidly improving nations are now known to include four from Latin America (Brazil, Chile, Mexico, and Peru) with only Argentina lagging, hardly surprising given its periodic demoralization over the last 30 years. Those who are sure that Hispanic migration to the United States will add a genetically determined low-IQ group to the population should think again. Also note that Rindermann and Thompson (above) used NAEP data to show that between 1971 and 2008, Hispanics reduced their gap with non-Hispanic whites from 11.59 to 8.46 IQ points. The 3.13 points gained is below black gains but the Hispanic community must absorb a continuous flow of recent immigrants.

9. Woodley, te Nijenhuis, and Murphy

Woodley, as we have seen, does not challenge the fact that large phenotypic IQ gains since the Victorian era have real world significance. He does not consider them "hollow" despite the fact that that they do not pass what I believe is a spurious test of their significance (whether they correlate with *g* loadings). The question is to what degree these have had to swamp counterproductive dysgenic trends, that is, the deterioration of genes for IQ caused by those of low IQ having more children than those of high IQ. In other words, to what degree have the causes of massive IQ gains had to add value to brains that were less and less receptive at conception?

They cite Richard Lynn's study of these trends but do not state his summary conclusion. Setting aside new additions to the gene pool from migration, Lynn puts the "genetic IQ" decline at about 4.4 points for the UK and 5.0 points for the US (Flynn, 2013, p. 44; Lynn, 2011, pp. 104 and 130). The UK total should be boosted to about 5.4 points because Lynn's estimate is only through 1980. This implies that social change over the 20th century was worth about 35 IQ points and that it had to overcome a resistance of 5 points for a net gain of 30. This 5-point loss is based on the only data I consider significant (for dysgenic trends see Flynn, 2013, pp. 39–58).

Lynn's estimate stands in stark contrast to the 14 points the authors derive from reaction time data.

The fact that IQ gains do not correlate with *g* is interpreted as indicating genetic decline. The point is moot in the sense that no one has been able to use this phenomenon as a measure of how great the genetic decline is supposed to be and, if it amounts to 5 points or less, it does not alter Lynn's estimate. Briefly, the case rests on the fact that *g* is thought to be significant in itself and correlates negatively with the effects of inbreeding depression and positively with heritability. Flynn (2013, pp. 15–19) attempts to clarify the true implications of these assertions.

The only significance of *g* in itself is that it is a measure of cognitive complexity. Therefore, the failure of IQ gains to tally with *g* is simply an indication that society did not select gains on various cognitive abilities to alter in accord with their degree of complexity (why should it: perhaps modernity selected less for what Vocabulary measures than for what map-reading measures). As for the correlation between cognitive complexity and the negative effect of inbreeding depression, this simply shows that the brain "areas" that deal with complexity are more fragile during sexual reproduction than other "areas" (they are more damaged by the effects of two recessives). Inbreeding influences IQ but the correlation with *g* is not significant. As for the correlation between cognitive complexity and heritability, this merely shows that family environment fades quicker for the cognitive abilities modernity has enhanced most. There are a host of environmental reasons why family environment might be more prolonged for say Arithmetic rather than Vocabulary.

The attempt to compare the per capita number of scientific geniuses in the Victorian era with today is not viable. First, the number of geniuses history records cannot be taken as an index of the genetic quality of the population at large. If so, the Scottish Enlightenment of the 18th century showed a huge advantage over England that was completely erased by the 19th century. Are we to believe that dysgenic reproduction had devastating effects for the Scots over a few generations? Do we really want to compare the average IQ of Greece and Rome in terms of the flowering of genius?

Second, when you reach the modern era, who do you classify as a genius? If you mean a really clever person who solves important problems, there are far more of them per capita today because far more people per capita are being educated to be scientists. I take an interest in the history of science and could list at least 100 "geniuses" since 1930. This is a 19th century hero worship term that designates people that were held in awe. Our capacity to accommodate a list of the awesome is psychologically limited. Imagine someone who claimed there were a thousand geniuses today, when none of us can even remember those who won the Nobel Prize. In any age, only a few can make the list and today there are so many contenders that hundreds are perforce ignored, as distinct from when Maxwell stood out against only a few competitors.

I consider Raphael Bousso a "genius". Ever heard of him? He has used the hypothesis that our universe is only one among an infinite and ever-growing assemblage of disconnected bubble universes to derive an estimate for dark energy, one that closely matches observations. His work holds out the prospect of a multi-universe theory that will generate novel predictions for our own bubble universe. It may even generate calculations that apply to the multi-universe (the collection of all possible universes) despite the fact that they are unobservable (lurk beyond our cosmic horizon). Now what if his hypothesis is falsified? He certainly will not be called a "genius" then, but the brilliant mind behind his work has nothing to do with whether the long-term data happen to confirm his theories.

This brings us to reaction times (RTs). It is taken to show a 14-point decrease in genic IQ between Galton's experiments circa 1889 and studies done as recently as 2004–2006. It is based primarily on data from Silverman (2010). The mere fact that we do not know whether Galton's participants did practice trials is unsettling, despite a quote from Johnson et al. (1985) to the effect "*As far we can ascertain* (italics mine), Galton took only a single measure each of visual and auditory reaction times" (p. 879). Galton found almost identical RTs for male and female contrary to all the literature existent. This is excused on the grounds that perhaps only an elite sample of women was willing to be tested.

I have no confidence about the history of RTs. Jensen believed that we could not compare even current RTs with one another because of the variability of the instruments (he was conducting a crusade to introduce standard equipment throughout the world). Ted Nettelbeck informs me that even since 1970, the methodology for measuring all elementary cognitive tasks has changed constantly rendering comparability impossible. The exception is Nettelbeck and Wilson (2004). Using the very same equipment, they found that between 1981 and 2001, Australian children made no Inspection Time (IT) gains but gained 5 points on the Peabody (PPVT).

It may be replied that elementary cognitive tasks would only pick up genetic trends. Lynn (2011, p. 169) argues that Australia had a strong dysgenic tendency through those born in 1965. The oldest IT Australian cohort was born about 1968. This poses the possibility of a conflict between dysgenic reproduction and ECT trends, unless of course dysgenic trends ceased circa 1965–68. This is possible. Flynn (2013, pp. 45–48) shows that they ended in Norway and Sweden well before that. There is also the vexed question of whether RTs really measure anything like neural speed (Flynn, 2009, pp. 69–74). If not, they cannot qualify as a measure of cognitive capacity over time. Flynn (1991) shows that they are conditioned by character differences between populations.

Reverting to the RT history Woodley et al. relate, even if one grants comparability, the internal evidence poses problems. Compare two eras: 1900 to 1943 (43 years) and 1943 to 2006 (63 years). I choose 1900 because it is alleged that RT studies at that time replicated Galton's results. To accommodate all studies (some had no female subjects), I have used males only for comparability. We go from 183.6 (1900) to 245 (average for 1943) to 236 (1970) to 223 (1992) to 315 (2000) and to 265.5 (2005). In sum, 61.4 of the 81.9 msec (75%) were lost before modern data became available. Multiplying 14 times 25% means that only 3.5 points were lost over the last 63 years. That rate is close to Lynn (4.4 points over 90 years).

10. Pietsching, Tran, and Voracek

This analysis of the Viennese clinic data takes me back many years (to 1986 when I first saw its results). Using scores

from the total WISC battery, Flynn (1987, p. 183) calculated IQ gains from 1962 to 1979. The new data come exclusively from a multiple choice vocabulary test administered from 1978 to 1994. To quote the author's summary comment: "The results of the IRT-based analyses do not provide strong and unequivocal support for changes in test-taking behavior, but instead suggest item drift of some subset of MWT-B items as a viable alternative explanation."

Item drift refers to the fact that some items that were difficult for subjects to answer correctly in 1978 became easier by 1994. This was particularly true of items that once baffled the lower half of the curve. The result was that the lower half of the curve made greater gains than the upper half, which reduced IQ variance. Test-taking behavior refers to the Brand hypothesis. He held that due to personality changes over time, subjects had become less responsible and were now "risk takers". Increasingly, they would not waste time reassuring themselves an answer was correct, would do more intelligent guessing, and would address all items (even if they were simply guessing) rather than leaving some items blank.

The fact that gains from 1978 to 1994 in Vienna predominated in the lower half of the curve does not alter the results of the total IQ data. Flynn (2012, pp. 41–43) sums up as follows. As for reduced variance over time, Belgium, Argentina, Sweden, Canada, and New Zealand do not show this. Israel does but only for females. Of seven nations for which we have data (mainly post-1948) from the whole curve, France, the Netherlands, America, and the UK (data lacking for older ages) show gains at all levels. Spain, Denmark, and Norway show the "bottom" phenomenon. Norway is very peculiar in that height gains predominate over the top half of the curve. Perhaps the upper classes took the advice of nutritionists about a healthy diet more seriously than the lower classes; and perhaps better schooling upgraded lower class education more than upper class education. Do note the significance of the latter: predominate gains over the lower half of the curve do *not* award nutrition some sort of victory over education as a factor. The literature is full of this assumption.

The choice of results from a multiple-choice vocabulary test rather than the WISC is crucial. Brand flatly asserted that his hypothesis did not apply to Wechsler tests. In some cases, the examiner makes sure that the subject simply cannot go on to more difficult items before ending performance on that subtest. Some subtests require subjects to verbally volunteer the correct answer (or solution) in a way that test sophistication would be unlikely to affect. For example, risk taking would be unlikely to help the subject put a city or nation in its proper locale, offer an adequate definition for a word like "delectable" (pseudo-example), or do mental arithmetic.

Brand's error was to ignore the huge gains on Wechsler tests around the world. Instead, he focused on the Scottish WISC, which he interpreted as showing small gains. In fact, by proceeding item by time, a method that ignored the low correlation between items, he underestimated the Scottish gains. The old Stanford–Binet method of taking test performance as a whole, and dividing mental age by chronological age, put the gains for two age groups at 15 to 17 points over 22.5 years. Correcting these by way of deviation IQs lowered the gains to about 12 to 13 points (Brand, 1990; Brand, Freshwater, & Dockrell, 1989; Flynn, 1990).

Brand confined his hypothesis to tests that had both multiple-choice items and a time limit. The Vienna test seems to set a generous time limit (if any). However, the subjects were instructed not to guess. With those directions, even subjects under no time pressure might evolve over 17 years. They might be transformed from those who left an answer blank because they were slightly doubtful about their response, into those who left it blank only if they were more doubtful. Thus, these results could be construed as casting doubt on Brand's hypothesis when test instructions maximized the likelihood of its predictions (Brand might interpret the instructions as ensuring that his hypothesis would have minimum applicability).

11. Must, O. and Must, A.

The above results came from Austria over the last generation. In that nation, test sophistication may have reached saturation point by 1979. Therefore, it does not obviate the Brand hypothesis for a nation whose subjects may have gone from zero test sophistication to much sophistication over the period in question.

Between 1930 and the present, Estonia probably illustrates the transition from "untested" to "test saturation". Must and Must found that penalizing wrong answers very much reduced Estonian IQ gains during that period (1933–36 and 2006), that is, they found a prominent role for test sophistication. However, as we saw above, when Shui et al. used IRT (which should have distinguished item drift from test sophistication), they found large Estonian gains on a range of IQ subtests. The contrast is odd.

As Must and Must note: "Students in modern societies have broad experience with different testing and examination procedures and their consequences. It is logical to assume that they are able to manage their test-taking resources just as they are able to manage different learning processes." This appeal to enhanced test sophistication explains their results without reference to the Brand hypothesis. By that I mean that better exam techniques (better time management and intelligent guessing) would occur purely as an adaption to both the growing frequency and significance of cognitive tests. No additional hypothesis about character altering from responsibility to irresponsible risk taking would be necessary.

Must and Must analyzed a test that was part multiple choice and part volunteered answers. It was not personally administered so no examiner encouraged them to do their best (offer definitions about which they were unsure). Therefore, whether a subject entered an answer into the space provided depended on self-confidence. In Must's introduction, there seems be an implication that IQ gains must tally with *g* to be cognitively significant. I will say a bit more about this below.

The format of Raven's Progressive Matrices is entirely multiple-choice and in principle, Brand's hypothesis might apply no matter whether the test is timed or untimed. In my opinion, Raven's gains are so huge that guessing could be only a minor factor. More important, evidence argues against it. More frequent "intelligent guessing" over time should make Raven's gains tail off as the practice grew towards saturation point. In The Netherlands, huge gains (20 points over a single generation) accelerated from 1952–62 to 1962–72 to 1972–82 (Flynn, 1987, p. 172). In Britain, Raven's gains were smaller prior to 1980 than thereafter (Flynn, 2012, pp. 43–49).

Raven's is really a kind of analogies test. Recently a brilliant article assessed what actually allows each generation to do better on it than the preceding generation (Fox & Mitchum, 2013). In 1900, long before Raven's was invented, Americans could do simple analogies grounded in the concrete world: domestic cats are to wild cats as dogs are to what? (Wolves). By 1961, they could handle two squares followed by a triangle implies two circles followed by what? (A semicircle: just as a triangle is half of a square, so a semicircle is half of a circle). By 2006, they could handle two circles followed by a semicircle implies two sixteens followed by what? (Eight: you have to see the relationship despite the transition from shapes to numbers).

Note how each step takes us further from the concrete world toward using logic on abstractions, eventually abstractions whose very identity shifts. These new habits of mind not only allow us to meet the demands of tertiary education and the professions but also upgrade our morality. We have gone from maxims (often cruel) that we simply inherit (like precious artifacts) to universalizing moral principles ever more inclusive and humane (Flynn, 2013, pp. 59–73).

12. Williams

I have left the overview to the end because some points have been covered in the specialized papers. I believe that a hierarchy of causes is at work.

The *ultimate cause* is the trend to modernity, caught but only partially so by the industrial revolution. A nation like Sudan has not industrialized much but may be influenced by visual culture via the mass media. The *intermediate causes* are the spinoffs of modernity, such as better nutrition, smaller family size, hothouse "education" of preschoolers, new parenting, more formal education, far more creative work roles and leisure, the new visual culture, and urbanization. Nations are at every stage on the road to modernity (or hardly on it at all) and the mix of these factors would vary dramatically at every stage. I endorse the comments about the multiplicity of factors. This does not forbid a modeling of the gains, but I suspect this would be subject to endless historical controversy (what nation at what time was where on the road to modernity) and elaboration (models multiplying as the process of modernity unfolded). I endorse the assertion that intermediate factors should lose potency as modernity runs its course. The *proximate causes* are what changes took place in the heads of people as they began to get more items correct on IQ tests. These causes were "diluted" by growing test sophistication at the start but eventually they emerged as people learned to use logic on abstractions, used abstractions to classify, took the hypothetical seriously, achieved larger vocabularies and funds of general information, became surrounded by visual imagery, and so forth.

I will not repeat my views on education versus nutrition, infant IQ gains, and top versus bottom of the curve. The paper's remarks about the absence of huge gains on crystallized (educationally loaded) tests are no longer valid: note the large gains American adults made on WAIS Vocabulary and Information discussed below.

My reservations are two. First, the absence of posited proximate causes, which are the key to a full explanation, and emerge only when you compare gains from one test to another and one subtest to another. For example, I believe a unique proximate explanation is needed for the huge Raven's gains (see Flynn, 2012, p. 57 for a list that shows how widespread and huge these are). Second the lack of a sociological dimension. I have no objection to Item Response Theory (IRT) as a measure of item shift in term of degree of difficulty. But even its use must be subject to sociological verification.

Let us imagine that IRT diminished the huge Vocabulary gain of American adults on the WAIS. Then imagine that TV programs between 1950 and 2006 showed that the vocabulary a mass audience was presumed to possess matched the gains. We could solve the dilemma by acknowledging that real-world vocabulary gains had occurred but that the WAIS failed to pick them up. Well it is logically possibly that cognitive abilities needed throughout the modern world (more vocabulary and general information, classification, taking the hypothetical seriously) were missed by tests specifically designed to measure them. But I do not find it plausible. We can sit in our studies and use our analytic tools as much as we wish, but nothing excuses the lack of real world social research.

13. The contributors taken together

I believe that a consensus is emerging that massive IQ gains over time have real-world consequences. Since 1900, Americans have gone from a people with a median of less than six years of schooling, through the mid-century high school revolution, into the tertiary revolution (12% of Americans had some tertiary experience in 1950 as compared to 52% today). Since 1900, Americans have gone from 3% in professional or sub-professional jobs to 35% in jobs that demand cognitive creativity, ranging from highly paid professionals (15%) to teachers, lower management, computer programmers and technicians (20%). It is inconceivable that their habits of mind did not alter, as the new educational and occupational demands and their cognitive abilities interacted with one another.

Flynn (2013) offers two new sets of data. First, between 1953–54 and 2006, there were huge adult Vocabulary gains on the WAIS, no doubt due to the expansion of tertiary education. These amounted to 17 IQ points. Perhaps it will be granted that vocabulary gains have social significance. During much the same period American children (who got no extra years of schooling) made minimal Vocabulary gains on the WISC. It is not likely that Americans were responsible children at school and turned into irresponsible risk takers as they matured, or that any other kind of test sophistication endowed them with the ability to offer definitions for difficult words. Adults gained 8 points on the Information subtest. Test sophistication would not endow them with knowledge about the locale of a city or nation.

Second, between 1995 and 2006, American adults made gains on the WAIS whose subtest magnitudes had a correlation with subtest *g* loadings somewhere between positive 0.540 and 0.621 (adjusted for restriction of range). During much the same period, American children made gains on the WISC whose subtest magnitudes had a correlation with subtest *g* loadings somewhere between negative 0.302 and 0.409. So one tallies with *g* and the other does not. I maintain that correlations with *g* are largely irrelevant to social significance. Today adults

communicate better and have wider general knowledge; children do not and have not.

Everyone concedes that people altered when the Enlightenment banished a mindset that tried animals in court and believed in witches. Did the alteration of our minds stop dead in 1900? We freed ourselves from fixation on the concrete and entered a world in which the mass of people began to use logic on abstractions and universalize their moral principles. Living our lives day by day, we take modernity for granted. The very existence of the modern world is astonishing. I refer not to the internet or the air travel or the organ transplants but to altered human beings and altered minds. Collectively the scholars in this volume are beginning to write the cognitive history of the 20th century.

References

Brand, C. R. (1990). A 'gross' underestimate of a 'massive' IQ rise? A rejoinder to Flynn. *Irish Journal of Psychology*, *11*, 52–56.

Brand, C. R., Freshwater, S., & Dockrell, W. B. (1989). Has there been a 'massive' rise in IQ levels in the West? Evidence from Scottish children. *Irish Journal of Psychology*, *10*, 388–394.

Cotton, S., Kiely, P., Crewther, D., Thompson, B., Laycock, R., & Crewther, S. (2005). A normative and reliability study of the Raven's Coloured Matrices for primary school aged children from Victoria, Australia. *Personality and Individual Differences*, *39*, 647–659.

Dickens, W. T., & Flynn, J. R. (2006). Black Americans reduce the racial IQ gap: Evidence from standardization samples. *Psychological Science*, *17*, 913–920.

Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *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.

Flynn, J. R. (1990). Massive IQ gains of the Scottish WISC: Evidence against Brand et al'.s hypothesis. *Irish Journal of Psychology*, *11*, 41–51.

Flynn, J. R. (1991). 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. (2008). *Where have all the liberals gone? Race, class, and ideals in America.* Cambridge University Press.

Flynn, J. R. (2009). *What is intelligence? Beyond the Flynn Effect [Expanded Paperback Edition].* Cambridge University Press.

Flynn, J. R. (2012). *Are we getting smarter: Rising IQ in the twenty-first century.* Cambridge University Press.

Flynn, J. R. (2013). *Intelligence and human progress: The story of what was hidden in our genes.* London: Elsevier.

Fox, M. C., & Mitchum, A. L. (2013). A knowledge based theory of rising scores on "culture free" tests. *Journal of Experimental Psychology: General*. http://dx.doi.org/10.1037/a0030155 (in press).

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

Johnson, R. C., McClearn, G., Yuen, S., Nagosha, C. T., Abern, F. M., & Cole, R. E. (1985). Galton's data a century later. *American Psychologist*, *40*, 875–892.

Lynn, R. (2011). *Dysgenics: Genetic deterioration in modern populations* (2nd ed.) Belfast: Ulster Institute for Social Research.

Nettelbeck, T. J., & Wilson, C. J. (2004). The Flynn effect: Smarter not faster. *Intelligence*, *32*, 85–93.

Silverman, I. W. (2010). Simple reaction time: It is not what it used to be. *The American Journal of Psychology*, *123*, 39–50.

Woodley, M. A., & Meisenberg, G. (2013). In the Netherlands the anti-Flynn effect is a Jensen effect. *Personality and Individual Differences*, *54*, 871–876.

The Flynn effect refers to a secular increase in population intelligence quotient (IQ) observed throughout the 20th century (1 â‡“ â‡“â€“4). The changes were rapid, with measured intelligence typically increasing around three IQ points per decade. The increase seemingly contradicted the earlier hypothesis that IQs were declining due to an inverse correlation between IQ and fertilityâ€”so-called dysgenic fertility (5). In recent years, the Flynn effect has weakened and reversed in several Western countries (6), leading to speculation that the Flynn effect was a transient phenomenon reflecting a boost in IQ from environmental factors that temporarily. The Flynn effect is the year-on-year rise of IQ test scores, an effect seen in most parts of the world, although at greatly varying rates. It was named by Richard Herrnstein and Charles Murray in The Bell Curve after the New Zealand based political scientist James R. Flynn, who did much to document it and promote awareness of its implications (Flynn, 1984, 1987). The average rate of rise seems to be around three IQ points per decade. Attempted explanations have included improved nutrition, a trend... The Flynn effect describes the tendency for IQ scores to rise across the board at a rate of approximately .30 points per year, or three points per decade (Flynn, 2009).Â Granting the aforementioned solutions to the â€˜Raven's paradoxâ€™, a question that arises is how uid cognition (Blair, 2006), as distinct from g as speed of information processing, e.g., â€˜chronometric gâ€™ (Jensen, 1998a, 2006), might be used in independently solving more â€˜domain specicâ€™ IQ-type problems (i.e., problems that tap test-specic rather than general sources.