

Human Varieties

PDF generated April 10, 2021 by Kalin's PDF Creation Station WordPress plugin

The Persistence of Cognitive Inequality: Reflections on Arthur Jensen's "Not Unreasonable Hypothesis" after Fifty Years

by Dalliard - Sunday, December 22, 2019

<https://humanvarieties.org/2019/12/22/the-persistence-of-cognitive-inequality-reflections-on-arthur-jensens-not-unreasonable-hypothesis-after-fifty-years/>

In 1969, *Harvard Educational Review* published a long, 122-page article under the title [“How Much Can We Boost IQ and Scholastic Achievement?”](#) It was authored by Arthur R. Jensen (1923–2012), a professor of educational psychology at the University of California, Berkeley. The article offered an overview of the measurement and determinants of cognitive ability and its relation to academic achievement, as well as a largely negative assessment of attempts to ameliorate intellectual and educational deficiencies through preschool and compensatory education programs. Jensen also made some suggestions on how to change educational systems to better accommodate students with disparate levels of ability.

While most of the article did not deal with race, Jensen did argue that it was “a not unreasonable hypothesis” that genetic differences between whites and blacks were an important cause of IQ and achievement gaps between the two races. This set off a huge academic controversy—Google Scholar [says](#) that the article was cited more than 1,200 times in the decade after its publication and almost 5,400 times by December 2019. The dispute about the article centered on the question of racial differences, which is understandable as Jensen’s thesis came out on the heels of the civil rights movement and its attendant controversies, such as school integration, busing of students, and affirmative action. Jensen questioned whether it is in fact possible to eliminate racial differences in socially valued outcomes through conventional policy measures, striking at the foundational assumption of liberal and radical racial politics. His floating of the racial-genetic hypothesis was what set his argument apart from the general tenor of the era’s scholarly and policy debate.

In this post, I will take a look at Jensen’s arguments and their development over time. The focus will be on the race question, but many related, more general topics will be discussed as well. The post has four parts. The first is a synopsis of Jensen’s argument as it was presented in the 1969 article. The second part offers an updated restatement of Jensen’s model of race and intelligence, while in the third part I argue, using the Bradford Hill criteria, that the model has many virtues as a causal explanation. In the fourth and concluding part I will make some more general remarks about the status and significance of racist thinking about race and IQ.[\[Note\]](#)

Contents

- [1. Synopsis](#)
 - [1.1. Failure of compensatory education](#)
 - [1.2. Measuring intelligence](#)
 - [1.3. IQ and occupational status](#)
 - [1.4. Intelligence, fixed or not?](#)
 - [1.5. General versus specific abilities](#)
 - [1.6. Population distribution](#)
 - [1.7. Decomposing the determinants of intelligence](#)
 - [1.8. Misconceptions about heritability](#)
 - [1.9. Empirical estimates of heritability](#)
 - [1.10. Effects of inbreeding](#)
 - [1.11. Heritability of special abilities](#)
 - [1.12. Heritability of scholastic achievement](#)
 - [1.13. Environmental effects](#)
 - [1.14. Physical and biological environment](#)
 - [1.15. Social class](#)
 - [1.16. Race differences](#)
 - [1.17. Raising intelligence](#)
 - [1.18. Dysgenics](#)
 - [1.19. Intensive educational interventions](#)
 - [1.20. Level I and II abilities](#)
 - [1.21. In a nutshell](#)
- [2. Updating Jensen's model of race and IQ](#)
 - [2.1. The development of Jensen's program](#)
 - [2.2. Modern restatement of Jensen's model](#)
 - [2.3.1. The model](#)
 - [2.3. Possible counterarguments to the default model](#)
 - [2.3.1. Race as a unit of analysis](#)
 - [2.3.2. Are hereditarians overtly egalitarian?](#)
 - [2.3.4. Modifiability of highly heritable characteristics](#)
 - [2.3.5. Flynn effect: The revolution that fizzled out](#)
 - [2.3.6. Realism run amok?](#)
- [3. Bradford Hill criteria and the black-white gap](#)
 - [3.1. Inferring causes](#)
 - [3.1.1. Strength \(effect size\)](#)
 - [3.1.2. Consistency \(reproducibility\)](#)
 - [3.1.3. Specificity](#)
 - [3.1.4. Temporality](#)
 - [3.1.5. Biological gradient](#)

- [3.1.6. Plausibility](#)
 - [3.1.7. Coherence](#)
 - [3.1.8. Experiment](#)
 - [3.1.9. Analogy](#)
 - [3.2. Conclusion](#)
- [4. The past, present, and future of race and IQ](#)
 - [4.1. Private truths](#)
 - [4.2. Who has the relevant expertise?](#)
 - [4.3. Why study it?](#)
- [Notes](#)
- [References](#)
- [Share this:](#)

1. Synopsis

Jensen's (1969a) article is often misrepresented, so I will start with a detailed synopsis of what he actually wrote. So as to not break the flow of the original argument, I have presented my comments, where pertinent, in footnotes, with a view to checking how Jensen's claims have stood the test of time.

1.1. Failure of compensatory education

Jensen started by noting that the Civil Rights Commission had concluded in its 1967 report to the Johnson administration that compensatory education programs designed to eliminate or reduce achievement gaps in schools had failed to do so. The report, which paid particular attention to high-quality programs in majority-black schools, found that "none of the programs appear to have raised significantly the achievement of participating pupils, as a group, within the period evaluated by the Commission."

Jensen argued that two theoretical ideas undergird preschool and compensatory education programs: the *average child concept* and the *social deprivation hypothesis*. The first of these refers to the idea that, save for a few rare individuals suffering from severe inborn neurological defects, cognitive abilities are pretty much the same in all children. Any variation between children is viewed as arising from unequal exposure to knowledge and skills before and outside of school. If such environmental inequality did not exist, all children would perform at more or less the same, adequate level in school.

The second and related idea is that the reason why racial minorities and children of poor parents tend to be below-average achievers in school is that they lack middle-class experiences that would give them the cognitive and non-cognitive skills needed for success. The purpose of preschool and compensatory education programs is then to provide “socially deprived” children with those crucial experiences.

1.2. Measuring intelligence

Next, Jensen embarked on a long discussion of the definition, measurement, and determinants of intelligence, which makes up the bulk of the article. In effect, the whole article is a critique of the average child concept and the social deprivation hypothesis.

Jensen wrote that it is not possible to unequivocally state what intelligence is. An operational approach is much more fruitful, and the “best we can do is to obtain measurements of certain kinds of behavior and look at their relationships to other phenomena and see if these relationships make any kind of sense and order.”[\[Note\]](#) The article then notes how influential the intelligence tests developed by Binet and Simon were, stressing that the Binet-Simon scales and their successors have their origins in the modern Western educational setting; abilities important for success in school were important when the domain of IQ items was defined. The correlation between IQ and scholastic achievement is about .5 to .6, but if data are aggregated longitudinally over many years, the correlation approaches the reliability of the measures.

Jensen then discussed the general factor of intelligence, or g , which was originally proposed by Charles Spearman at the beginning of the 20th century. Performance on all cognitive tests is positively correlated, and the strengths of the test-test correlations cannot be accounted for by superficial (dis)similarities in test content or format. Spearman argued that the core of general intelligence was “the ability to educe relations and correlates.” Similarly, Thomas Aquinas defined intelligence as “the ability to combine and separate,” or “to see the difference between things which seem similar and to see the similarities between things which seem different.” The best measures of g require such abilities.

Jensen noted that attempts to create tests of complex problem solving that do not give rise to a general factor have failed. The g factor accounts for 50% or more of the total individual differences variance in a typical test battery.[\[Note\]](#) Jensen argued that if the term intelligence is used, it should refer to g .

It has been found that a test consisting of tasks designed by Jean Piaget and his collaborators for the study of mental growth in children is loaded on the general factor about as highly as psychometric tests are. Therefore it “seems evident that what we call general intelligence can be manifested in many different forms and thus permits measurement by a wide variety of techniques.” Jensen pointed to cross-modal transfer as a central characteristic of intelligence—this means, for example, the recognition of the same stimulus when it is administered in a different

sensory modality, such as tactile versus visual representation.

The g loading of a test will vary as a function of the nature of the tests together with which it is factor-analyzed. Jensen argued that it is possible to fractionate g into smaller sources of variance. He analogized g with a general athletic ability factor that might be derived from a test consisting of various athletic performances.[\[Note\]](#)

The article endorsed Raymond Cattell's dichotomy of fluid and crystallized intelligence. Fluid intelligence is the capacity for new learning and problem solving, whereas crystallized intelligence represents previously acquired skills and knowledge. Among people sharing a common culture, the two are expected to be highly correlated. Fluid intelligence may top out as early as late teens, whereas crystallized intelligence is expected to accumulate up to old age.

1.3. IQ and occupational status

According to Jensen, IQ is related to occupational status mainly through educational attainment. The *average* educational and income levels associated with occupations, the prestige of occupations, and the mean IQs of occupations are all correlated with each other at around the .80–.90 range. At the individual level, the correlation between occupational status and IQ has been found to be between .42 and .71 in various studies. Within occupations, the correlation between IQ and job proficiency has been found to be around .20–.25. The correlation between IQ and ease of training for various occupational skills appeared to be around .50.[\[Note\]](#)

1.4. Intelligence, fixed or not?

Jensen criticized the frequently used term “fixed intelligence” as a misnomer that confuses genotypes and phenotypes. Genotypic influence on intelligence is fixed in the sense that the genetic factors are fully laid down at conception. Intelligence, however, is properly a phenotype and it is never fixed, because it is “a result of the organism's internal genetic mechanisms established at conception and all the physical and social influences that impinge on the organism throughout the course of its development.” The interesting question is the correlation between genotypes and phenotypes at various points in development. The square of this correlation is known as *heritability*.

The article looked at the evidence for the stability of IQ, viz., the extent to which individuals retain their standing relative to others when retested over the course of time. IQ, like other developmental characteristics, is rather unstable in small children but becomes increasingly stable throughout childhood. After age 8, the stability correlation, when corrected for measurement error, is between .9 and unity.[\[Note\]](#)

1.5. General versus specific abilities

Jensen argued that the term intelligence should be reserved for the general factor, or *g*. On the one hand, he emphasized that *g* or IQ tests do not capture the full extent of mental abilities. On the other hand, he stressed that *g* reflected a biological reality and captured capabilities that have been “singled out as especially important by the educational and occupational demands prevailing in all industrial societies.”

1.6. Population distribution

Jensen went on to discuss the distribution of IQ. Originally IQ scores were defined in relation to mental age, but, due to certain problems with the mental age concept, modern tests define IQ in relation to scores obtained in a same-aged norming sample. IQ scores are normally distributed by construction, but Jensen argued that there are reasons to believe that normality is not just a convenient assumption. Firstly, he referred to the central limit theorem and the expectation that the sum scores of a large set of test responses would be normally distributed, with the caveat that IQ items do not fully match the premises of the theorem, e.g., they are not uncorrelated. Secondly, Jensen argued that if the normal distribution of IQ scores is warranted, IQ scores should behave like an interval scale. Evidence for this has been obtained from sibling studies that show the regression of sibling IQs on one another to be linear almost throughout the IQ scale, with the same mean number of IQ points separating siblings regardless of IQ level. The exception to this pattern (and the normality of IQ scores) is found at the low end of the scale where there is an excess of scores as a result of genetic and chromosomal abnormalities and other pathological conditions. The IQ distribution therefore resembles that of height which is also normally distributed (within sexes) with a “bulge” at the lower end due to dwarfism.

Low IQ due to pathology, sometimes called *organic mental retardation*, is usually accompanied by abnormal physical appearance, and must be distinguished from *familial mental retardation*, the latter referring to low IQ scores that are part of the normal distribution of IQ. Evidence for the validity of this distinction comes from sibling comparisons. It has been found that there is no correlation between the IQs of severely mentally retarded individuals and their siblings, whereas in the mildly retarded range we find the usual sibling correlation. Further evidence for the distinction is provided by the observation that severe mental retardation occurs at similar rates across all social classes, whereas mild retardation is concentrated in the lower classes. There is some indication that the density of IQ scores at the high end of the scale is greater than the normal distribution predicts, too, but the evidence for this is inconclusive.[\[Note\]](#)

1.7. Decomposing the determinants of intelligence

Next, Jensen discussed the inheritance of cognitive ability, with the aim of countering the “belief in the almost infinite plasticity of intellect, the ostrichlike denial of biological factors in individual differences.” He argued that “the slighting of the role of genetics in the study of intelligence can only hinder investigation and understanding of the conditions, processes, and

The article discussed selective breeding studies in animals, paying specific attention to those that bred rats for greater learning ability. As to the applicability of breeding schemes to humans, Jensen quoted a 1967 position statement by three eminent geneticists (James F. Crow, James V. Neel, and Curt Stern) who had written that a “selection program to increase human intelligence (or whatever is measured by various kinds of ‘intelligence’ tests) would almost certainly be successful in some measure. The same is probably true for other behavioral traits. The rate of increase would be somewhat unpredictable, but there is little doubt that there would be progress.”^[Note]

The article presented the following decomposition of intelligence into various components:

where: V_P = phenotypic variance in the population
 V_G = genic (or additive) variance
 V_{AM} = variance due to assortative mating. $V_{AM} = 0$ under random mating (panmixia).
 V_D = dominance deviation variance
 V_I = epistasis (interaction among genes at 2 or more loci)
 V_E = environmental variance
 Cov_{HE} = covariance of heredity and environment
 V_I = true statistical interaction of genetic and environmental factors
 V_e = error of measurement (unreliability).

8 / 106

Jensen pointed out that assortative mating can have a substantial effect on the population distribution of IQ:[\[Note\]](#)

[A]ssortative mating increases the genetic variance in the population. By itself this will not affect the mean of the trait in the population, but it will have a great effect on the proportion of the population falling in the upper and lower tails of the distribution. Under present conditions, with an assortative mating coefficient of about .60, the standard deviation of IQs is 15 points. If assortative mating for intelligence were reduced to zero, the standard deviation of IQs would fall to 12.9. The consequences of this reduction in the standard deviation would be most evident at the extremes of the intelligence distribution. For example, assuming a normal distribution of IQs and the present standard deviation of 15, the frequency (per million) of persons above IQ 130 is 22,750. Without assortative mating the frequency of IQs over 130 would fall to 9,900, or only 43.5 percent of the present frequency. For IQs above 145, the frequency (per million) is 1,350 and with no assortative mating would fall to 241, or 17.9 percent of the present frequency. And there are now approximately 20 times as many persons above an IQ of 160 as we would find if there were no assortative mating for intelligence. Thus differences in assortative mating can have a profound effect on a people's intellectual resources, especially at the levels of intelligence required for complex problem solving, invention, and scientific and technological innovation.

On genotype-environment correlations Jensen wrote:

A genotype for superior ability may cause the social environment to foster the ability, as when parents perceive unusual responsiveness to music in one of their children and therefore provide more opportunities for listening, music lessons, encouragement to practice, and so on. A bright child may also create a more intellectually stimulating environment for himself in terms of the kinds of activities that engage his interest and energy. And the social rewards that come to the individual who excels in some activity reinforce its further development. Thus the covariance term for any given trait will be affected to a significant degree by the kinds of behavioral propensities the culture rewards or punishes, encourages or discourages. For traits viewed as desirable in our culture, such as intelligence, hereditary and environmental factors will be positively correlated. But for some other traits which are generally viewed as socially undesirable, hereditary and environmental influences may be negatively correlated.

[...]

In making overall estimates of the proportions of variance attributable to hereditary and

environmental factors, there is some question as to whether the covariance component should be included on the side of heredity or environment. But there can be no “correct” answer to this question. To the degree that the individual’s genetic propensities cause him to fashion his own environment, given the opportunity, the covariance (or some part of it) can be justifiably regarded as part of the total heritability of the trait. But if one wishes to estimate what the heritability of the trait would be under artificial conditions in which there is absolutely no freedom for variation in individuals’ utilization of their environment, then the covariance term should be included on the side of environment. Since most estimates of the heritability of intelligence are intended to reflect the existing state of affairs, they usually include the covariance in the proportion of variance due to heredity.

Regarding gene-environment interactions Jensen noted:

There is considerable confusion concerning the meaning of interaction in much of the literature on heredity and intelligence. It is claimed, for example, that nothing can be said about the relative importance of heredity and environment because intelligence is the result of the “interaction” of these influences and therefore their independent effects cannot be estimated. This is simply false. The proportion of the population variance due to genetic \times environment interaction is conceptually and empirically separable from other variance components, and its independent contribution to the total variance can be known.

1.8. Misconceptions about heritability

Next, Jensen discussed certain common conceptual errors in debates about heritability. He noted that while both environments and genes are needed for an organism to exist at all, that does not render the question of the relative importance of nature and nurture meaningless. The relevant question concerns the proportions of population variation that can be attributed to each.

While noting that heritability is a population statistic that cannot be used to partition “a given individual’s IQ into hereditary and environmental components”, Jensen wrote that because heritability is the squared correlation between phenotypes and genotypes, it can be used to make probabilistic statements “concerning the average amount of difference between individuals’ obtained IQs and the ‘genotypic value’ of their intelligence.”

The article also noted that heritability is not a constant but depends on the amount of variability in

the causal factors. Increasing causally relevant environmental variation will decrease heritability, while increasing genetic variation (while holding the environment constant) will increase heritability.

Responding to the criticism that heritability estimates are meaningless because we cannot “really” measure intelligence, Jensen pointed out that the estimates show that whatever the tests measure, it is heritable. To the extent that the tests are not “culture-free”, heritability estimates will be decreased.

Another common counterargument is that we must be able to “spell out in detail every single link in the chain of causality from genes (or DNA molecules) to test scores if we are to say anything about the heritability of intelligence.” According to Jensen, this is not so because “[s]elective breeding was practiced fruitfully for centuries before anything at all was known of chromosomes and genes, and the science of quantitative genetics upon which the estimation of heritability depends has proven its value independently of advances in biochemical and physiological genetics.”

Still another conceptual confusion is that because something like one’s vocabulary cannot be directly inherited, it cannot be heritable. But people differ markedly in “the amount, rate, and kinds of learning they evince even given equal opportunities.” High heritability indicates that opportunities for learning have been widespread, while low heritability shows the opposite.

High heritability does not necessarily imply immutability. Large changes in environmental conditions may change heritability, or heritability may remain the same while the population mean changes. However, Jensen argued that the degree of heritability says something about “the locus of control of a characteristic”:

The control of highly heritable characteristics is usually in the organism’s internal biochemical mechanisms. Traits of low heritability are usually controlled by external environmental factors. No amount of psychotherapy, tutoring, or other psychological intervention will elicit normal performance from a child who is mentally retarded because of phenylketonuria (PKU), a recessive genetic defect of metabolism which results in brain damage. Yet a child who has inherited the genes for PKU can grow up normally if his diet is controlled to eliminate certain proteins which contain phenylalanine. Knowledge of the genetic and metabolic basis of this condition in recent years has saved many children from mental retardation.

Finally, Jensen pointed out that children share only half of the genetic variants of each parent (or somewhat more under assortative mating). This means that substantial phenotypic differences

between parents and children (or between siblings) are compatible with high heritability.

1.9. Empirical estimates of heritability

For empirical estimates of the heritability of IQ, Jensen relied particularly on various publications by Cyril Burt based on British data^{[\[Note\]](#)}, and on Erlenmeyer-Kimling and Jarvik's (1963) review of kinship correlations for IQ which included data from 58 studies across eight countries.

Jensen summarized Burt's report of a variance decomposition of Stanford-Binet IQs based on many types of kinship pairs drawn primarily from London schools. This analysis found a heritability of 81% (for unadjusted test scores) or 93% (for test scores adjusted after retesting children whose IQs did not match their teachers' impressions of their "brightness"). *These results must be regarded with caution given that serious concerns about the authenticity of Burt's data were raised some years after Jensen's article was published.*

Next, Jensen looked at Erlenmeyer-Kimling and Jarvik's data on IQ correlations among many types of kinship pairs. These data were based on various tests and different testing conditions, and were collected by "numerous investigators with contrasting views regarding the importance of heredity." Nevertheless, the results were generally compatible with the view that there is a strong polygenic component to IQ differences.

Supplementing Erlenmeyer-Kimling and Jarvik's data with some of Burt's, Jensen presented the following table where empirically observed correlations are compared to correlations from two theoretical, purely genetic models:

TABLE 2

Correlations for Intellectual Ability: Obtained and Theoretical Values

<i>Correlations Between</i>	<i>Number of Studies</i>	<i>Obtained Median r^*</i>	<i>Theoretical Value¹</i>	<i>Theoretical Value²</i>
<i>Unrelated Persons</i>				
Children reared apart	4	— .01	.00	.00
Foster parent and child	3	+ .20	.00	.00
Children reared together	5	+ .24	.00	.00
<i>Collaterals</i>				
Second Cousins	1	+ .16	+ .14	+ .063
First Cousins	3	+ .26	+ .18	+ .125
Uncle (or aunt) and nephew (or niece)	1	+ .34	+ .31	+ .25
Siblings, reared apart	33	+ .47	+ .52	+ .50
Siblings, reared together	36	+ .55	+ .52	+ .50
Dizygotic twins, different sex	9	+ .49	+ .50	+ .50
Dizygotic twins, same sex	11	+ .56	+ .54	+ .50
Monozygotic twins, reared apart	4	+ .75	+ 1.00	+ 1.00
Monozygotic twins, reared together	14	+ .87	+ 1.00	+ 1.00
<i>Direct Line</i>				
Grandparent and grandchild	3	+ .27	+ .31	+ .25
Parent (as adult) and child	13	+ .50	+ .49	+ .50
Parent (as child) and child	1	+ .56	+ .49	+ .50

* Correlations not corrected for attenuation (unreliability).

¹ Assuming assortative mating and partial dominance.

² Assuming random mating and only additive genes, i.e., the simplest possible polygenic model.

The table indicates that there are some systematic departures from theoretical expectations, presumably reflecting non-genetic influences. Jensen illustrated these departures in the following graph where the median correlations reported by Erlenmeyer-Kimling and Jarvik are shown for different kinds of kinship pairs reared together and apart:

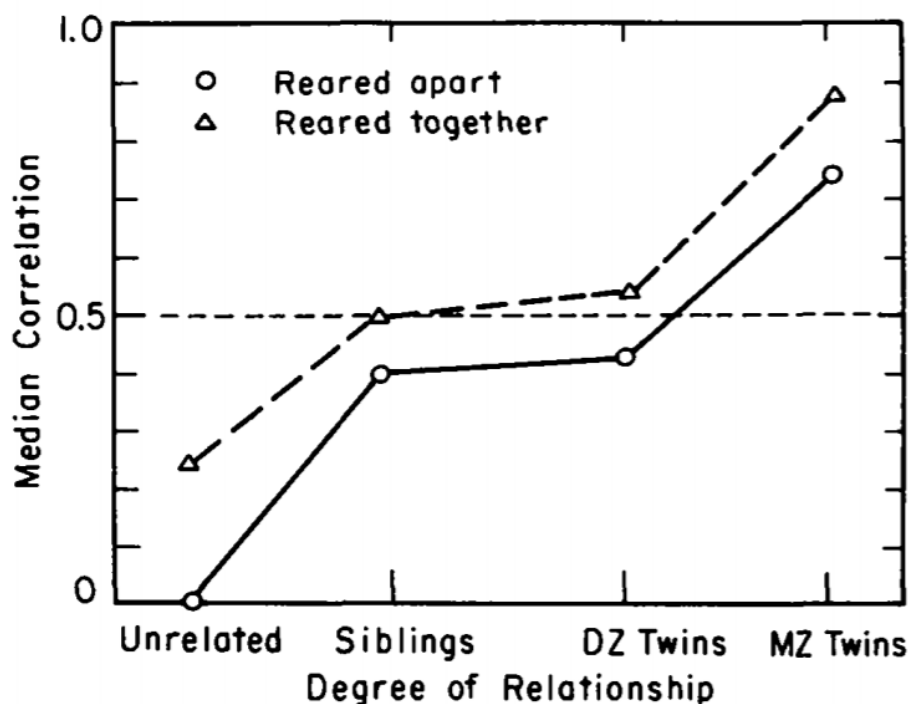


FIGURE 6.

Median values of all correlations reported in the literature up to 1963 for the indicated kinships. (After Erlenmeyer-Kimling & Jarvik, 1963.) Note consistency of difference in correlations for relatives reared together and reared apart.

Jensen's own calculations, originally reported in a previous paper (Jensen, 1967) and drawing on Erlenmeyer-Kimling and Jarvik's correlations for monozygotic and dizygotic twins, suggested an average heritability of 80% for IQ, after adjustment for test unreliability. His estimate for the shared environmental effect was 12%, with 8% left for the unshared environment. Adding Burt's

data to the mix, Jensen arrived at an estimate of 77%, which rose to 81% after adjustment for unreliability, which he regarded as the best overall estimate of the heritability of IQ that data allowed at the time.

The correlation between monozygotic twins reared apart provides the conceptually simplest estimate of heritability, provided that environments are uncorrelated within the pairs. Jensen had access to three studies of MZ twin pairs reared apart. Newman, Freeman, and Holzinger's (1937) study of 19 pairs found a correlation of .77 (.81 corrected for unreliability), and the correlation in Shields' study of 44 pairs was .77 (.81 corrected), too. Burt (1966) reported a study of 53 pairs where the correlation was .86 (.91 corrected). Jensen regarded Burt's study as the best of the bunch because it had the largest and most representative sample, along with very early separation of the twins.[\[Note\]](#)

Correlations between children and their foster versus biological parents provide another way to estimate the effects of nature and nurture. Jensen reported that correlations for IQ between adoptive children and their foster parents were between 0 and .20, while correlations between children and their biological parents "gradually increase from zero at 18 months of age to an asymptotic value close to .50 between ages 5 and 6", and "this is true whether the child is reared by his parents or not."

Still another way to gauge the effects of heredity and the environment is to correlate adoptive children's IQs with measures of their rearing environment. On this score, Jensen paid particular attention to the early study by Burks (1928), where the relevant multiple correlation was .42, suggesting that the measured environment accounted for 18% of IQ variance. Burks calculated that the average effect of a one standard deviation change along the environmental scale was six IQ points, which, Jensen noted, was half the magnitude of the average IQ difference between ordinary siblings reared together. Burks also collected data on parents raising their own biological children, along with measures of rearing environments. Sewall Wright later used Burks' data to arrive at an estimate of 81% for the heritability of IQ (Wright, 1931).[\[Note\]](#)

1.10. Effects of inbreeding

The article went on to discuss research on the effect of inbreeding on IQ. A negative effect is expected because inbreeding increases the probability that an individual will have two defective mutant recessives in a given locus. A large study conducted in Japan after the Second World War found that the children of cousins averaged almost eight points lower than the children of unrelated parents after controlling for age and socioeconomic status. Additionally, an American study found a high incidence of mental retardation among children from nuclear incest matings (brother-sister or father-daughter).

1.11. Heritability of special abilities

The article reported that the heritabilities of non-*g* cognitive abilities have been found to range from near zero to about .75, with most values between .50 and .70. Several broad abilities have genetic variance independently of *g*. Few studies had investigated the heritability of non-cognitive skills at the time. However, some evidence on motor skill learning suggested that heritability may be even higher in the non-cognitive sphere than for intelligence.

1.12. Heritability of scholastic achievement

Jensen argued that the heritability of scholastic achievement was considerably less than that of intelligence. He calculated an average heritability of 40% for a variety of scholastic measures, while the contributions of shared and unshared environmental components were 54% and 6%, respectively. He found that estimates for the heritability of scholastic achievement varied over a much wider range than those for IQ, with lower estimates in primary school and for simple forms of learning and somewhat higher estimates in high school and for more complex forms of learning. He referred specifically to twin data from the National Merit Scholarship Corporation in which about 60% of the variation in class ranks could be attributed to the shared family environment. This, together with the small within-family environmental variance, pointed to the family environment exerting a strong influence on scholastic performance. Unrelated individuals reared together were also found to be much more similar in school performance than IQ.[\[Note\]](#)

Jensen argued that environmentally malleable non-cognitive skills were important determinants of scholastic achievement. He proposed that efforts to improve school performance by improving non-cognitive skills had a much better chance of success than efforts to improve intelligence:[\[Note\]](#)

Thus it seems likely that if compensatory education programs are to have a beneficial effect on achievement, it will be through their influence on motivation, values, and other environmentally conditioned habits that play an important part in scholastic performance, rather than through any marked direct influence on intelligence per se. The proper evaluation of such programs should therefore be sought in their effects on actual scholastic performance rather than in how much they raise the child's IQ.

1.13. Environmental effects

Jensen argued that the effect of the environment on IQ is non-linear. Moving children from an extremely deprived environment to a normal environment can boost IQ by up to dozens of points, but children reared in average environments do not get an appreciable boost from being placed in a cognitively enriched environment. Below a certain threshold of environmental quality, deprivation can have a large negative effect on IQ, but above the threshold environmental variations are mostly inconsequential. The vast majority of participants in research on the

heritability of IQ are drawn from environments above the threshold, which accounts, in part, for the finding of high heritability for IQ.

Jensen compared the effect of the environment on IQ to that of nutrition on height. Nutritional deficiencies will lead to stunting but above a certain level of nutritional adequacy, even great variations in eating habits have little effect on stature.

Jensen set the threshold of environmental deprivation under which IQ is strongly affected at a low level. He did not think that a mere lack of middle-class amenities would have a great effect on IQ. To have a large effect, the environment must impose severe sensory and motor restrictions on the child. Examples of such deprivation include mentally retarded orphanage children studied by Skeels & Dye's (1939), and the case of "Isabel", a girl who was confined to an attic and reared by a deaf-mute mother until age 6. Children brought up in such conditions have been found to recover and attain normal intelligence after they are placed in a normal environment.

According to Jensen, children described as "culturally disadvantaged" do not generally encounter severe environmental deprivation.[\[Note\]](#) If severely deprived children are moved to an adequate environment, they usually attain large and permanent IQ gains, while the IQ gains of culturally disadvantaged children placed in an enriched environment are slight and transitory. In contrast to severely deprived children, culturally disadvantaged children do not show early cognitive deficits, and they experience average or, sometimes, precocious perceptual and motor development. As culturally disadvantaged children grow up, their IQs become more strongly correlated with those of their parents, which points to genetic influence. On the other hand, less intelligent parents may also be less able to provide their children with environmental conditions that are conducive to intellectual development.

The article then discussed a longitudinal study by Heber et al. (1968) which investigated a sample of children born to poor black mothers in Milwaukee, WI. The mean IQ of children of mothers with sub-80 IQs declined over time, while the mean IQ of children of mothers with above-80 IQs did not show a temporal trend.[\[Note\]](#) Jensen suggested that this is not consistent with environmental deprivation, and is more consistent with genetic factors exerting greater influence in older children. He also discussed an old study by Wheeler (1942) of "Tennessee mountain children" where two cohorts of children aged between 6 and 16 were tested in 1930 and 1940. The average IQ of the 1940 cohort was 10 points higher, presumably due to environmental improvements, but both cohorts showed a similar decline in norm-referenced IQ from age 6 to 16.[\[Note\]](#)

Next, Jensen discussed the concept of reaction range according to which "similar genotypes may result in quite different phenotypes depending on the favorableness of the environment for the development of the characteristic in question" and "some genotypes may be much more buffered against environmental influences than others." Different genetic strains may therefore have

dissimilar heritabilities for a given trait.

Jensen emphasized that heritability estimates represent average heritabilities, and that heritabilities may therefore differ between subpopulations within the same population. He noted that all major studies of the heritability of IQ are based on samples of white people, with no studies of blacks, for example. Reaction norms may vary between races in that “some genetic strains may be more buffered from environmental influences”, which would suggest that heritabilities may not be the same for different populations even in the very same environment.[\[Note\]](#) The availability of heritability estimates for different populations would, however, be useful for the purposes of testing some hypotheses regarding genetic and environmental causes of group differences in IQ.

1.14. Physical and biological environment

Jensen put the environmental variance of IQ scores at around 20%. He noted that people tend to think that the environmental variance is due to differences in “social and interpersonal environment, child rearing practices, and differences in educational and cultural opportunities afforded by socioeconomic status.” However, Jensen argued, much or even most of the environmental variance may not be associated with social factors but rather with certain physical and biological environmental factors, with the implication that “advances in medicine, nutrition, prenatal care, and obstetrics” are important for the improvement of intelligence.

The first piece of evidence for the importance of the non-social environment that Jensen discussed is the fact that twins scored an average of 4 to 7 points lower in IQ tests than singletons.[\[Note\]](#) These differences were likely due to prenatal factors as it seems unlikely that twins and singletons would receive such disparate postnatal treatment, especially as the twin deficit had been observed across social classes. The fact that “MZ twins have a higher mortality rate and greater disparity in birth weights than DZ twins” suggested that “MZ twins enjoy less equal and less optimal intrauterine conditions than DZ twins or singletons.” Boy twins averaged lower IQs than girl twins, which is consistent with the general observation that male infants are more vulnerable to prenatal impairment. Birthweight is modestly correlated with later IQ independently of sociocultural factors, and in MZ twin pairs the twin who weighs less at birth usually has a lower IQ in school age. This may be because “the unequal sharing of nutrients and space stunts one twin more than its mate.” Much of IQ variation in MZ twins therefore appears to be due to prenatal environmental factors. The significance of this observation is that differences in intrauterine conditions between singletons in the general population may contribute substantially to IQ variation.

Next, Jensen discussed certain medical techniques claimed to improve the intrauterine environment. He noted that more optimal intrauterine and perinatal conditions appear to be associated with precocious perceptual-motor development. This is a conundrum for those who argue that inadequate prenatal care and complications of pregnancy are to blame for the lower

mean IQ of blacks, for black infants do not typically exhibit subnormal development.

Disadvantageous reproduction-related factors and conditions, such as pregnancies at early ages and in close succession, low birth weight, prematurity, and infant mortality, are correlated with race and social class, but it is unclear to what extent they can account for IQ gaps. 75–80% of cases of mental retardation cannot be explained by known complications of pregnancy, brain damage, or gene and chromosomal defects, and therefore presumably represent the low end of the normal polygenic distribution of IQ. Research indicates that when common reproductive difficulties occur singly, they have no effect on the child's intellectual status, suggesting that “the nervous system is sufficiently homeostatic to withstand certain unfavorable conditions if they occur singly.”

Reviewing the literature on the effect of premature birth on IQ, Jensen noted that prematurity has a strong relation to brain dysfunction but that the crucial factor appears to not be prematurity per se but low birth-weight. The latter seems to act as “a threshold variable with respect to intellectual impairment.” The incidence of babies weighing less than 5.5 lb is greater in lower social classes, but socioeconomic variables do not account for more than 1% of the total variance in birth-weight.

Black babies weigh less, on average, than white babies even after controlling for social class, but they also mature at a lower birth-weight than white babies. Prematurity and low birth-weight are more common in blacks than whites, but this is not a full explanation of the black-white IQ gap because black children perform significantly less well in cognitive tests than white children matched for birth-weight.

If race differences are ignored and prematurity is defined as a condition where birth-weight is less than 5.5 lb, the association between prematurity and lower IQ can be statistically explained by the common factor of social class. Social class does not, however, fully explain the association between low IQ and prematurity for birth-weights less than 3 lb. The association between IQ and low birth-weight is mainly due to cases of severe mental retardation among very low-weight infants, and otherwise the association is very weak by school age. It is also possible that there are individual differences in genetic predisposition for prenatal impairment.

Severe undernutrition in early years results in lower IQ. If it occurs after early childhood, it appears to have no permanent effect. For example, severely malnourished prisoners of war have been found to suffer no intellectual deficiencies once returned to normal living conditions. Extreme undernutrition is rare in the United States, but some unknown proportion of the urban population might benefit from nutritional supplementation.

First-borns have been shown to have higher IQs, on average, than later-borns. The reason for this phenomenon is probably biological rather than social-psychological. Jensen asserted that it is “almost certainly not a genetic effect.” The birth-order advantage is slight, and is conspicuously

observed only in the extreme right tail of the distribution of achievement.[\[Note\]](#)

1.15. Social class

An extensive literature from many countries shows that children's IQs are associated with the socioeconomic status (SES) of their parents.[\[Note\]](#) The correlation is typically in the range of .35–.40. Jensen pointed out that this correlation is almost a logical necessity because IQ is heritable and the educational system and the occupational hierarchy act as (imperfect) intellectual screening processes.

Jensen cast doubt on the idea of SES as an important *cause* of IQ by noting that the IQs of children reared apart from their siblings and parents are correlated with the IQs and educational and occupational levels of such biological relatives, and the correlations are almost as strong as among intact families. Moreover, among siblings raised together, those with IQs above the family average tend to move up the SES scale as adults, while those with IQs below the family average tend to move down.

There is a negative correlation between SES and Developmental Quotient in children under age 2, while there is an increasing positive correlation after age 2. Low-SES children thus get a “head-start” on development, but this trend is reversed at later ages when tests become less motoric and more loaded on the general factor of intelligence.

1.16. Race differences

Moving to a more focused discussion of race differences, Jensen started by outlining his social philosophy that puts the individual at the center of the picture:

The variables of social class, race, and national origin are correlated so imperfectly with any of the valid criteria on which the above decisions should depend, or, for that matter, with any behavioral characteristic, that these background factors are irrelevant as a basis for dealing with individuals—as students, as employees, as neighbors. Furthermore, since, as far as we know, the full range of human talents is represented in all the major races of man and in all socioeconomic levels, it is unjust to allow the mere fact of an individual's racial or social background to affect the treatment accorded to him. All persons rightfully must be regarded on the basis of their individual qualities and merits, and all social, educational, and economic institutions must have built into them the mechanisms for insuring and maximizing the treatment of persons according to their individual behavior.

Jensen noted that if people considered social problems only from the perspective of individuals, there would be no “race problem.” That is, however, a philosophy that few adopt. People like to compare groups, and assess whether there are groups differences in “the most desirable and the least desirable social and occupational roles in a society.” The fact that different races are disproportionately represented in such roles in America, and that so much current thinking revolves around this fact compels research on all the reasons why this inequality exists. To what extent is the inequality due to unfairness, or the use of intrinsically irrelevant criteria like skin color by decision-makers, and to what extent is it due racial differences in the distributions of indisputably relevant characteristics? According to Jensen, these questions can be answered “only through unfettered research”, and no reasonable hypothesis must be ruled out of court for ideological reasons. Attitudes to the contrary “represent a danger to free inquiry and, consequently, in the long run, work to the disadvantage of society’s general welfare.”

Everyone agrees, Jensen wrote, that environmental causes, including past history, contribute to intellectual, educational, and occupational disparities between whites and blacks in America. However, the possible contribution of heredity to racial differences has been “greatly ignored, almost to the point of being a tabooed subject, just as were the topics of venereal disease and birth control a generation or so ago.”

Groups that are geographically or socially isolated from each other for many generations will differ in their gene pools, and will therefore probably show differences in highly heritable phenotypic traits. Races are “breeding populations” where matings are much more common within-population than between populations. Technically, races are distinguished by their different distributions of allele frequencies. Genetic differences are manifested in “virtually every anatomical, physiological, and biochemical comparison one can make between representative samples”, and this surely applies to the brain as well.

The pertinent question about racial-genetic differences in behavioral traits is not about their existence but about the direction and magnitude of the differences, and the medical, social, educational, and other consequences thereof. Some genetic differences are of no consequence, and the idea that all genetic differences arise and persist due to natural selection cannot be accepted.

Dreger and Miller (1960, 1968) and Shuey (1966) reviewed the evidence for black-white IQ differences. Whites outscore blacks by 15 points (1 standard deviation) on average, and this magnitude is quite similar across 81 different tests analyzed in Shuey’s review of 382 studies. 15 percent of blacks outscore the average white individual. If black and white populations were of the same size, 23 percent of IQ differences would be explained by race, with within-race differences explaining 77 percent. Controlling for SES reduces the black-white gap to about 11 points. Blacks perform relatively worse in tests that are “culture-free” or “culture-fair”, in tests of abstract abilities, and in non-verbal tests. The variance of black IQ scores has been found to be lower than that of whites—by 40% according to one study.

In tests of scholastic achievement, whites and Asian Americans outscore blacks by about 1 standard deviation, as indicated by the [Coleman Report](#). The gap is relatively constant from grades 1 through 12. Puerto Ricans, Mexican-Americans, and American Indians outscore blacks to a smaller degree.

The black-white disadvantage cannot be completely or directly explained by discrimination or inequitable schooling. Given that intelligence variation is strongly influenced by genetics, it is not unreasonable to propose that genetic differences may be involved in the black-white gap. While this hypothesis has been met with forceful condemnation in social science, “it has been neither contradicted nor discredited by evidence.”

Jensen formulated his position on the plausibility of the genetic explanation in the following way:

The fact that a reasonable hypothesis has not been rigorously proved does not mean that it should be summarily dismissed. It only means that we need more appropriate research for putting it to the test. I believe such definitive research is entirely possible but has not yet been done. So all we are left with are various lines of evidence, no one of which is definitive alone, but which, viewed all together, make it a not unreasonable hypothesis that genetic factors are strongly implicated in the average Negro-white intelligence difference. The preponderance of the evidence is, in my opinion, less consistent with a strictly environmental hypothesis than with a genetic hypothesis, which, of course, does not exclude the influence of environment or its interaction with genetic factors.

Jensen then went on to enumerate various points that he viewed as especially relevant for understanding the causes of the black-white IQ gap. He noted that no one has managed to show that controlling statistically for environment and education would equalize black and white IQs. Then he drew attention to the fact that the proportion of mentally retarded children (defined as $IQ < 75$) is higher in black families than white families at all SES levels. While an environmental hypothesis would predict less of a difference at high SES levels, that is not in fact observed. A genetic hypothesis supplies a ready explanation for this observation: regression to the mean.

Research shows, in fact, that low-SES white children outscore high-SES blacks, on average, and it seems improbable that the cultural opportunities available to poor white children would be superior to those available to middle- and upper-class black children. While environmental explanations of the regression effect have been devised, they often seem to strain credibility.

Environmental explanations of group differences are usually *ad hoc* in nature: they provide a plausible explanation for the particular case that they were devised to explain but lack generality across situations. The existence of an environmental difference is never a sufficient causal

explanation of a group difference. Jensen used the example of father absence as an explanation of the black-white IQ gap, pointing out that research does not support the idea that father's presence or absence makes an independent contribution to IQ or scholastic outcomes.

The Coleman Report assessed many socioeconomic and environmental factors often believed to be major sources of individual and group differences in scholastic performance, such as reading material and cultural amenities in the home, structural integrity of the home, foreign language in the home, preschool attendance, parents' education, parents' educational desires for child, parents' interest in child's school work, time spent on homework, and child's self-concept (self-esteem). These factors were all correlated with scholastic performance within races and ethnic groups, but they were not systematically related to group differences. For example, American Indians were more disadvantaged when it came to environmental factors than blacks, yet they outscored blacks in ability and achievement tests.[\[Note\]](#)

Black infants have been found to develop precociously, especially motorically, when compared to white infants. Developmental precocity correlates negatively with parental SES in whites. High-SES black infants are more precocious than high-SES white infants, while no difference has been found between low-SES black and white infants. These findings can be considered in light of the fact that adverse prenatal, perinatal, and postnatal complications lead to developmental delay. Black precocity in comparison with whites is also found for certain physiological indices of development, such as the rate of ossification of cartilage. Black babies also mature at a lower birth-weight than white babies.

In adults, the largest sampling of white and black IQ scores comes from the administration of the the Armed Forces Qualification Test (AFQT) to representative samples of millions of men. As of 1966, the overall failure rates for whites and blacks in the test were 19 percent and 68 percent, respectively, with an eligibility cut-off point equivalent to an IQ of 86 or so. Approximately half of the black families were middle-class and above when the AFQT was administered, so "even if we assumed that all of the lower 50 percent of Negroes on the SES scale failed the AFQT, it would still mean that at least 36 percent of the middle SES Negroes failed the test, a failure rate almost twice as high as that of the white population for all levels of SES."[\[Note\]](#)

Jensen ended his discussion of race differences here, asking, perhaps rhetorically, whether such findings question the credibility of exclusively environmental explanations of observed differences.

1.17. Raising intelligence

Jensen noted that the cognitive demands of work are rising, and there will be increasingly fewer jobs available for low-IQ individuals. Thus the advantages of raising intelligence seem obvious. He noted that a small cognitive elite, perhaps 2 percent of the population, is probably responsible for civilizational advances, but that the rest are able to assimilate and enjoy the consequences of

the advances. He also noted that tests and degrees can become barriers to entry to jobs where high cognitive ability is correlated with performance but is not required. He suggested that making people more adept at the “essential requirements of a given job” is a more feasible goal than raising intelligence or academic achievement.

Jensen wrote that that given the high heritability of IQ and the threshold nature of environmental effects on it, solely improving the environment of the economically disadvantaged would not be expected to lead to large IQ gains. He criticized an argument by Milton Schwebel, according to whom providing adequate environmental conditions for the currently environmentally deprived children (estimated at 26 percent of the population) would boost their IQs by 20 points. Jensen noted that this is unrealistic considering that it would boost the IQs of the deprived above those of the non-deprived already enjoying adequate environments.

Jensen argued that not only educational services but also public health, social services, and welfare and employment practices are important for boosting intelligence. However, he warned that direct improvements in the environment could have indirect biological consequences as well.

1.18. Dysgenics

Jensen reviewed data on fertility and IQ, concluding that white fertility does not appear to be dysgenic, but that black fertility appears to be so.[\[Note\]](#) On the latter finding, he wrote:

Is there a danger that current welfare policies, unaided by eugenic foresight, could lead to the genetic enslavement of a substantial segment of our population? The possible consequences of our failure seriously to study these questions may well be viewed by future generations as our society's greatest injustice to Negro Americans.

1.19. Intensive educational interventions

Jensen noted that while large-scale compensatory programs have made little difference for the disadvantaged, much more positive results have been obtained from intensive small-scale experiments “where maximum cultural enrichment and instructional ingenuity are lavished on a small group of children by a team of experts.”

Small enrichment and cognitive stimulation programs have been found to result in gains of around 5–20 points for IQ, and around 0.5–2 standard deviations for scholastic achievement. An analysis by Rick Heber of 29 intensive preschool programs found an average gain of 5 to 10 IQ points at the end of preschool. More gains in IQ are seen when the program comprises special cognitive training. The most intensive programs go beyond the classroom and involve daily

sessions in the child's home, and some programs of this sort have reported gains of up to 20 IQ points. Gains appear to be restricted to children from deprived backgrounds, and are not seen in non-disadvantaged children.

Next, Jensen cast doubt on the findings of intensive intervention studies by pointing to certain limitations in them. These include:

- Lack of control groups in many studies.
- Selection of kids for programs on the basis of low IQ, which means that they tend to achieve gains simply due to the regression to the mean. Studies with control groups almost always show a regression effect in the control group.
- Relative ease of achieving IQ gains in small children—for example, getting two additional Stanford-Binet items right boosts IQ from 85 to 93 at age four, while at age 10 the boost is from 85 to 88.
- Materials similar to those used in IQ tests are often found in nursery schools, which may explain IQ gains. Jensen recounted once visiting “an experimental preschool using the Stanford-Binet to assess pretest—post-test gains, in which some of the Stanford-Binet test materials were openly accessible to the children throughout their time in the school as part of the enrichment paraphernalia.”[\[Note\]](#)
- Pre-intervention IQ scores are often poorly measured because small children from deprived backgrounds are unfamiliar with the testing situation. Gains during preschool programs may be gains in test-savviness rather than ability.
- It is not clear what the psychometric nature of intervention gains in IQ is—are they on g, or on something less important?[\[Note\]](#)
- It is notable that gains from enrichment programs are of a similar magnitude and durability as the effects of direct coaching and practice on IQ.
- The fadeout effect has been widely observed. Gains in IQ due to enrichment programs have a strong tendency to vanish over time.[\[Note\]](#)
- Dubious scalability of small-scale interventions.
- Enrichment programs often work by precipitating the acquisition of skills and knowledge that children would acquire anyway at a somewhat later age. Children can learn many things “prematurely” by associative/rote learning, but that does not mean that complex cognitive structures are being developed.

Next, Jensen reviewed some prominent intervention studies:

- The Indiana Project involved deprived, low-IQ Appalachian white children five years of age. A special year-long kindergarten program resulted in gains of 4–10.8 IQ points when compared to control groups.
- The Perry Preschool Project in Ypsilanti, Michigan, involved disadvantaged, low IQ children and their parents, seeking to remedy especially verbal skills. There was a gain of

8.9 IQ points after one year of the preschool, but by the end of second grade the gain had faded to 1.6 IQ points, a non-significant difference from the control group.

- The Early Training Project run from the Peabody College was an enrichment and cognitive stimulation program that involved disadvantaged children and their mothers. Four years after the start of the program, the experimental group had gained 7.2 IQ points over a control group.
- The Durham Education Improvement Program was a preschool program for children from impoverished homes. The participants attained average gains of 2.62 to 9.27 IQ points, depending on the test.
- The Bereiter-Engelmann program at the University of Illinois sought to teach specific cognitive skills and scholastic knowledge in small groups. Gains over 18 months were 8–10 IQ points and higher for specific content tests. No control group was used. According to Jensen, the program showed that scholastic performance is easier to boost than IQ, at least in the early years.
- A preschool program by Merle Karnes at the University of Illinois attempted to ameliorate specific learning deficits in disadvantaged 3-year-olds. By age 4 the experimental group gained 19.7 IQ points over the control group. A small sample size and lack of longer-term follow-up preclude strong interpretations of these findings.

According to Rosenthal and Jacobson's famous Pygmalion experiment, low expectations by teachers explain why disadvantaged children perform relatively poorly on IQ tests. They tested the effect of manipulating teachers' expectations on the IQ development of randomly chosen students, and found that it significantly boosted IQs. Jensen noted that there were various deficiencies in the design of the experiment, and in how the results have been analyzed, and suggested that it be "replicated under better conditions before any conclusions from the study be taken seriously or used as a basis for educational policy." [\[Note\]](#)

Jensen tentative overall conclusion regarding attempts to increase IQ was that the payoff from preschool and compensatory education programs is small. He thought that improving scholastic performance was a more feasible goal and one that should be emphasized in lieu of attempts to boost IQ. Educators should also assess gains using tests of specific skills rather than IQ. Jensen thought that raising general intelligence was more in the province of the biological sciences than psychology and education. He also thought that the goal of making disadvantaged children indistinguishable from middle-class children was unrealistic.

1.20. Level I and II abilities

Jensen postulated the existence of two types of learning ability: Associative learning ability (Level I), and cognitive or conceptual learning and problem-solving ability (Level II). Level I abilities involve relatively little transformation of the input, and there is a high correspondence between the stimulus input and the response output. Digit memory, serial rote learning, and recall of visually or verbally presented materials are some tasks that are thought to tap into Level I

abilities. Level II abilities, in contrast, involve transformation and elaboration of the stimulus input to construct a response. Tests with a high g loading and a low cultural loading, such as Raven's matrices, tap into Level II abilities.

The significance of the Level I/Level II distinction is that while there are large racial and SES differences in Level II abilities (favoring white and middle- and upper-class individuals), group differences in Level I abilities are small. Jensen therefore believed that the distinction was of a "great potential importance to the education of many of the children called disadvantaged." Because traditional classroom instruction was principally developed to be compatible with the ability patterns of middle-class students, schools place a greater emphasis on cognitive learning than associative learning. To maximize the potential of students from different genetic and cultural backgrounds, associative learning should be given a greater role in teaching.

While Jensen thought that more research was needed to fully substantiate the Level I/Level II taxonomy^[Note], he nevertheless regarded it as a highly promising way of thinking about learning, and closed the article by proposing that schooling should not be uniform but rather reflect the diversity of human abilities:

If diversity of mental abilities, as of most other human characteristics, is a basic fact of nature, as the evidence indicates, and if the ideal of universal education is to be successfully pursued, it seems a reasonable conclusion that schools and society must provide a range and diversity of educational methods, programs, and goals, and of occupational opportunities, just as wide as the range of human abilities. Accordingly, the ideal of equality of educational opportunity should not be interpreted as uniformity of facilities, instructional techniques, and educational aims for all children. Diversity rather than uniformity of approaches and aims would seem to be the key to making education rewarding for children of different patterns of ability. The reality of individual differences thus need not mean educational rewards for some children and frustration and defeat for others.

1.21. In a nutshell

Much of Jensen's article was an explication of basic theories and findings from psychometrics and quantitative genetics. Given that the foundational ideas and many excellent empirical studies in those fields date to the first half of the 20th century, Jensen's article was built on a firm foundation, and there are many things in the article that can still be read with profit. His critical comments on what can and cannot be learned from experiments seeking to boost intelligence and educational achievement remain germane, too.

Jensen's answer to the two-part question posed in the title of his paper was that there appears to be not much that can be done to boost IQ, but that scholastic achievement is much more amenable to intervention. The latter conviction stemmed from the fact that, firstly, much more than IQ is involved in school success, and, secondly, that those non-IQ things are either reasonably evenly distributed across races and social classes (Level I abilities), or strongly influenced by the shared environment and therefore presumably more modifiable. With the benefit of hindsight, both of the claims about school achievement seem overly optimistic. As detailed in the notes to the synopsis, Jensen had a misplaced faith in the importance of what he called Level I abilities, and his estimate of the contribution of the shared environment to academic achievement appears much too high in light of later research.

As Johnson (2012) pointed out much later, Jensen's preparation for his article was so thorough that he anticipated just about all the criticisms that were to be presented against his thesis. The attacks on the article rarely came from perspectives that he had failed to consider. It was more a question of how much importance and credibility Jensen versus his critics attributed to a given viewpoint. Nevertheless, much less was known about intelligence and differences between races in those days, and the article was only a starting point for Jensen's later work.

2. Updating Jensen's model of race and IQ

2.1. The development of Jensen's program

It was a stroke of good fortune for race realism as a paradigm and research program that Arthur Jensen turned his considerable talents to questions of intelligence and group differences. Looking at his early publications from the 1950s through the mid-1960s, it was not at all obvious that he was to have a defining influence on the study of intelligence and race. Early on, he published little or nothing on these topics, and his research was focused on the experimental study of human learning, and clinical and personality psychology. His psychological education at Berkeley and Columbia had been behaviorist and psychoanalytic in theoretical orientation, and it was only through self-study and personal contacts with individual differences researchers and geneticists that he was eventually able to escape those then fashionable intellectual dead-ends.^[Note] He pointed to his sojourn in Hans Eysenck's lab in London as a turning point in his career, reflecting on its importance in the following manner decades later:

I emphasize my postdoctoral work with Eysenck, because I believe it planted the seeds of virtually everything I have done since then. It put me on the path that I have followed, in one way or another, for all of my later research. Although each of the many subsequent byways could not have been anticipated, they all led more or less consistently in one general direction—what came to be known as the London School of differential psychology, originated by Galton and with Spearman, Burt, and Eysenck successively as its leading exponents. (I knew personally only Eysenck and Burt.) The London School is

not really a school or even a doctrine or a theory. Rather, it is a general view of psychology as a *natural science* and as essentially a branch of *biology*. Its central concern is variability in human behavior. It is *Darwinian* in that it views both *interspecies* variation and an important part of *intraspecies* variation (both individual and group differences) in certain classes of behavior as products of the evolutionary process. It is *behavior-genetic* in that the evolutionary process depends upon genetic variation and selection, and the neural basis of behavioral capacities is subject to these evolutionary mechanisms the same as other physical characteristics. It is *quantitative* in that it emphasizes the objective measurement and taxonomy of behavior and the operational definition of latent traits or hypothetical constructs. It is *analytical* in that it subjects quantitative data to mathematical formulation and statistical inference. It is *experimental* in that it typically obtains measurements, both behavioral and physiological, under specifically defined and controlled conditions. It is *reductionist* in that it aims theoretically to explain complex phenomena in terms of simpler, more elemental processes. It is *monistic* (as opposed to dualistic) in that it neither posits nor seeks any explanatory principle that does not consist of strictly physical processes: it views complex psychological phenomena as emerging solely from interactions among more elemental neurophysiological processes and their past and present interactions with environmental conditions.

–Jensen, 1998a

In an early article reviewing theories of personality, such as they stood in the 1950s, Jensen pointed out that “theories in psychology are seldom disproved; they just fade away” (Jensen, 1958). This ephemerality and non-incrementalism make much of research in psychology frivolous. In contrast to that, through Jensen’s work race realism became anchored in some of the strongest, most permanent ideas in the social and behavioral sciences, ones that have not shown signs of fading away over the last 100 or more years.

What started as a provisional hypothesis about race and IQ in the 1969 paper grew, over decades, into a full-blown research program. Partly to fill in lacunae in the research literature and partly to respond to criticisms of his arguments, Jensen ended up studying almost every aspect of the problem with characteristic thoroughness. An early outline of this research program is presented in his 1973 book *Educability and Group Differences*. This book, which is still worth reading, is an update and extension of the 1969 article, and is both more tightly argued and more confident and expansive in its outlook. It contains at least an inkling of everything Jensen was to pursue in his research until his death in 2012.

The next milestone among his publications was *Bias in Mental Testing*, a 1980 tome that was

instrumental in putting to rest the once popular conjecture that test bias is an important explanation of the black-white IQ gap. In 1985, Jensen published “The nature of the black–white difference on various psychometric tests: Spearman’s hypothesis”, a *Behavioral and Brain Sciences* target article where he laid out the case that the general factor of intelligence, or *g*, was the main locus of black-white IQ differences (Jensen, 1985). As I will outline later, Spearman’s hypothesis is important because of the specificity that it confers to the IQ gap.

The elaboration of the meaning and significance of the *g* construct became a great preoccupation of Jensen’s career. While the question of the structure of cognitive ability is logically separate from the question of race differences, Jensen’s monistic view of differences within and between groups meant that his research interests fed fruitfully into each other. He formulated his monism in this way:

There is fundamentally, in my opinion, no difference, psychologically and genetically, between individual differences and group differences. Individual differences often simply get tabulated so as to show up as group differences—between schools in different neighborhoods, between different racial groups, between cities and regions. They then become a political and ideological, not just a psychological, matter.

—Jensen (1973)

According to Spearman’s hypothesis, the nature of *g* and the nature of race differences are intertwined. In 1998, Jensen published his magnum opus, *The g Factor: The Science of Mental Ability*. This book remains the go-to source for anyone wishing to understand cognitive ability. The book, especially its chapters 11 and 12, contains Jensen’s fullest statement of his research program on race differences in IQ. He wrote about what he termed the *default hypothesis* in this way (p. 444):

The default hypothesis states that human *individual* differences and *population* differences in heritable behavioral capacities, as products of the evolutionary process in the distant past, are essentially composed of the same stuff, so to speak, controlled by differences in allele frequencies, and that differences in allele frequencies *between* populations exist for all heritable characteristics, physical or behavioral, in which we find individual differences *within* populations.

With respect to the brain and its heritable behavioral correlates, the default hypothesis holds that individual differences and population differences do not result from differences in the brain’s basic structural operating mechanisms per se, but result entirely from other aspects of cerebral physiology that modify the sensitivity, efficiency, and effectiveness of

the basic information processes that mediate the individual's responses to certain aspects of the environment.

[...]

The population differences reflect differences in allele frequencies of the same genes that cause individual differences. Population differences also reflect environmental effects, as do individual differences, and these may differ in frequency between populations, as do allele frequencies.

Jensen denied that there are essential, qualitative differences in the cognitive abilities of blacks and whites. Rather, he stated that the differences are [accidental](#) (in the philosophical sense) and quantitative. Smart blacks and smart whites and stupid blacks and stupid whites are smart and stupid in the same ways, and black-white differences are of the same nature as the differences between smart and stupid subgroups of whites or blacks. The between-race differences are due to the same causes, genetic and environmental, that operate within each race.[\[Note\]](#)

2.2. Modern restatement of Jensen's model

Using CFA terminology (for definitions, see e.g., Brown, 2014), Jensen's *default model*, or my interpretation thereof, has the following properties:

1. A latent, reflective *g* factor is the largest source of variance and covariance in intellectual tasks among individuals.
2. The black-white IQ gap is around 1 standard deviation, favoring whites, and is mainly (>>50 percent) a *g* factor gap. There are black-white differences in non-*g* factors and test specificities as well, but these are minor compared to the *g* gap and sometimes directionally favor blacks. "Black" and "white" refer to self-identified groups.
3. The *g* factor can be measured without bias in blacks and whites using commonly administered IQ tests. In other words, *strict measurement invariance* with respect to race is usually attainable.
4. Individual differences in IQ in both blacks and whites are predominantly (up to 80% in adults) due to additive genetic influences.
5. Differences between whites and blacks in IQ scores are mainly (>50 percent) due to differences in the frequencies of alleles influencing *g*. The same alleles influence both within-race and between-race differences.
6. Some (<50 percent) of the IQ gap may be due to non-genetic influences, including microenvironmental effects (e.g., lead poisoning) and indirect genetic effects ("genetic nurture"), especially when IQ is measured in childhood. Sociological and social-

psychological causes, such as discrimination, have at best a minor effect on the gap.

7. The model properly concerns the black-white IQ gap in the United States only, but to the extent that white and black Americans can be considered as representative samples of the indigenous populations of Europe and sub-Saharan Africa, the model has much to say about the rest of the world as well.

This list is, firstly, a general conceptual representation of the default model, and all empirical analyses of the black-white gap can be thought of as tests of this model, even if many of the listed properties are not necessarily explicitly considered in such analyses. Secondly, it can be considered as a description of a single structural equation model (SEM) that at least potentially incorporates all the listed properties in a single analysis. I will discuss the possible identifiability (i.e., whether its parameters can be estimated from data) of such a model below.

Theoretical impetus for this model derives, besides Jensen's work, especially from an important paper by Lubke et al. (2003) where it is argued that psychometric measurement invariance has strong causal implications in the study of group differences: when measurement invariance holds, the factors that explain differences within groups must also explain them between groups; the sources of group differences are the same as the sources of individual differences within each group. Given that strict measurement invariance for IQ appears to usually hold between black and white Americans (e.g., Dolan, 2000; Dolan & Hamaker, 2001; Lubke et al., 2003; Trundt, 2013; Frisby & Beaujean, 2015; Lasker et al., 2019), the multigroup measurement invariance model can be seen as a phenotypic manifestation of Jensen's default hypothesis.

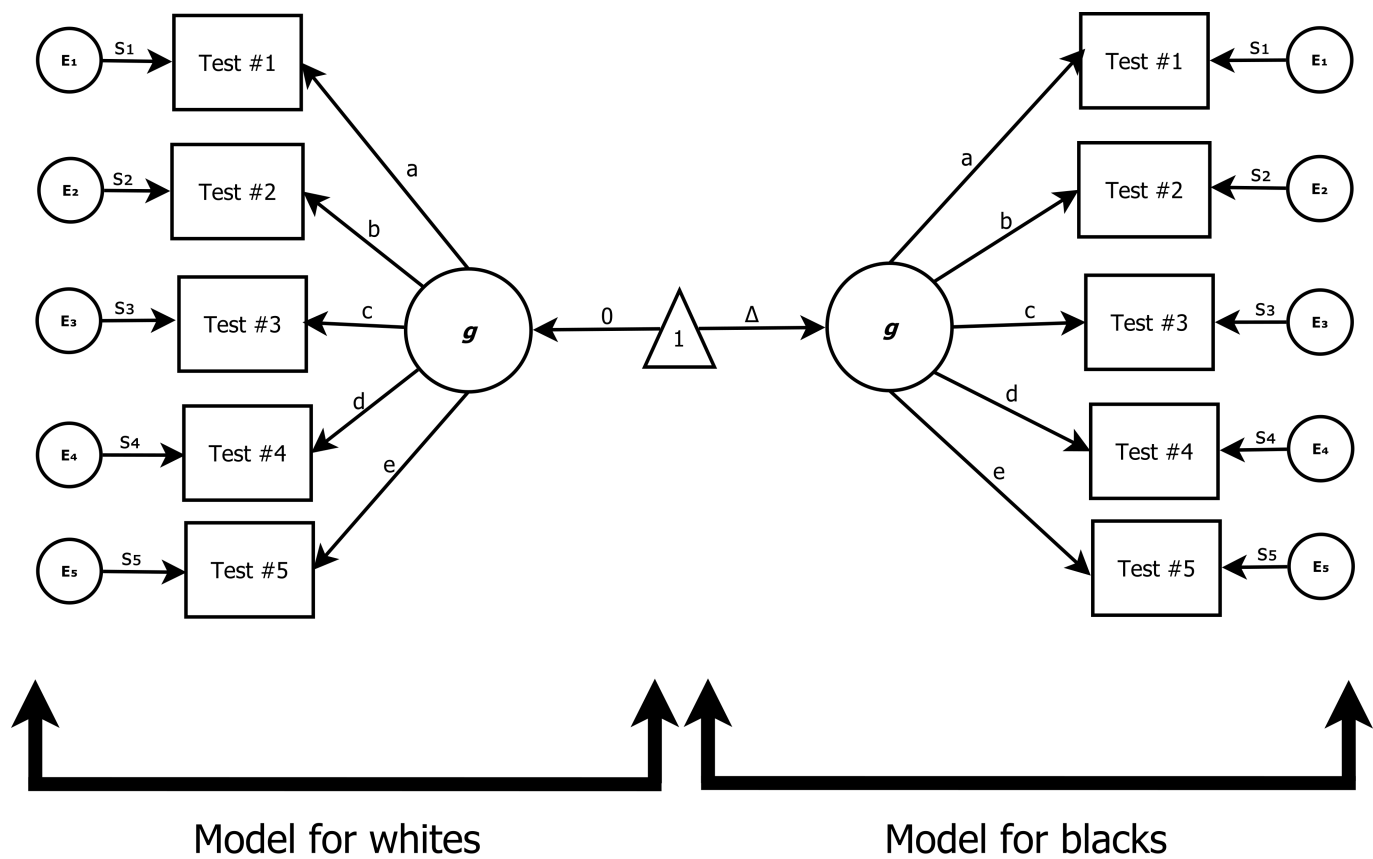
Given the causal implications of the measurement invariance model, a study finding that invariantly measured latent factors and g in particular are highly heritable in both blacks and whites would shift the weight of evidence, buttressing the default model. It would not, however, be a direct test of the model. To directly investigate the etiology of race differences, we need to combine psychometric models with biometric ones. Most biometric models deal only with variances and covariances (or correlations) among observed variables, while the means of the variables, and thus group differences, are ignored. However, as shown in Dolan et al. (1992), it is possible to incorporate means in biometric models, for example in the classical twin model.

A practical example of biometric multigroup modeling is Rowe and Cleveland's (1996) study of academic achievement in a sample of black and white children. This study—which found evidence of a genetic contribution to black-white differences—is an important proof of concept, but the fact that the sample used was small and unrepresentative, and consisted of pairs of full and half siblings (rather than MZ and DZ twin pairs which can be analyzed in a more powerful and better understood framework) limits its credibility. Another limitation is that the study was based on observed test scores, making it vulnerable to the criticism that the results simply recapitulate the biases of the tests used.

2.3.1. The model

Black-white IQ differences can be depicted as an SEM model that incorporates (most of) the properties enumerated at the beginning of the previous section. While the biometric part of the model appears to not be a practical tool at the moment, I will discuss its properties because it is of theoretical interest and because I think that it holds much promise and could be further developed.

The model can be divided into psychometric and biometric stages. The first, psychometric stage would look like this (click for a larger image):



To clarify my notation, circles are latent variables to be estimated and rectangles are observed variables (tests), while the triangle denotes a mean difference between whites and blacks (the “1” in the triangle is simply a notational convention). The parameters of the model are indicated with letters and numbers on the paths that connect the variables. The arrowheads show the direction of causality for each association between the variables in the model. For notational simplicity, it is assumed that the variances of all variables are unity.

There are five tests in the model, and their covariances are accounted for by the latent g factor. The residuals of the tests (e_1 – e_5) not explained by g represent measurement error and variance specific only to a particular test. The model is formally a multigroup confirmatory factor model, with the white submodel on the left and the black submodel is on the right.

The first stage of the model can also be called the strict measurement invariance model, because its purpose is to establish that strict invariance holds between whites and blacks. An analysis of measurement invariance involves establishing that the following four conditions are met in a multigroup confirmatory factor model (note that each condition is a precondition for the next one on the list):

1. *Configural invariance*: The number of common factors is the same across groups, and the same indicators (tests) load on the same factors in all groups.
2. *Metric invariance*: The factor loadings are the same across groups.
3. *Scalar invariance*: The intercepts of the indicators are equal across groups, i.e., mean differences on indicators between groups are consistent with the size of the factor loadings and can be attributed to the latent common factor(s).
4. *Strict invariance*: Variances of the residuals are equal across groups.

If these conditions are met in a sequential testing scheme where various indices of fit can be used, differences between groups can be attributed to the same latent factor (or factors) that causes differences within groups. See Lubke et al. (2003) for more discussion of measurement invariance and group differences. I have discussed the topic [here](#), too.

It can be seen that exactly the same variables and paths are included in the black and white submodels. This indicates that configural invariance is expected to be true. The factor loadings a – e are denoted with the same letters in blacks and whites, which means that they are constrained to be equal across races. This corresponds to metric invariance.

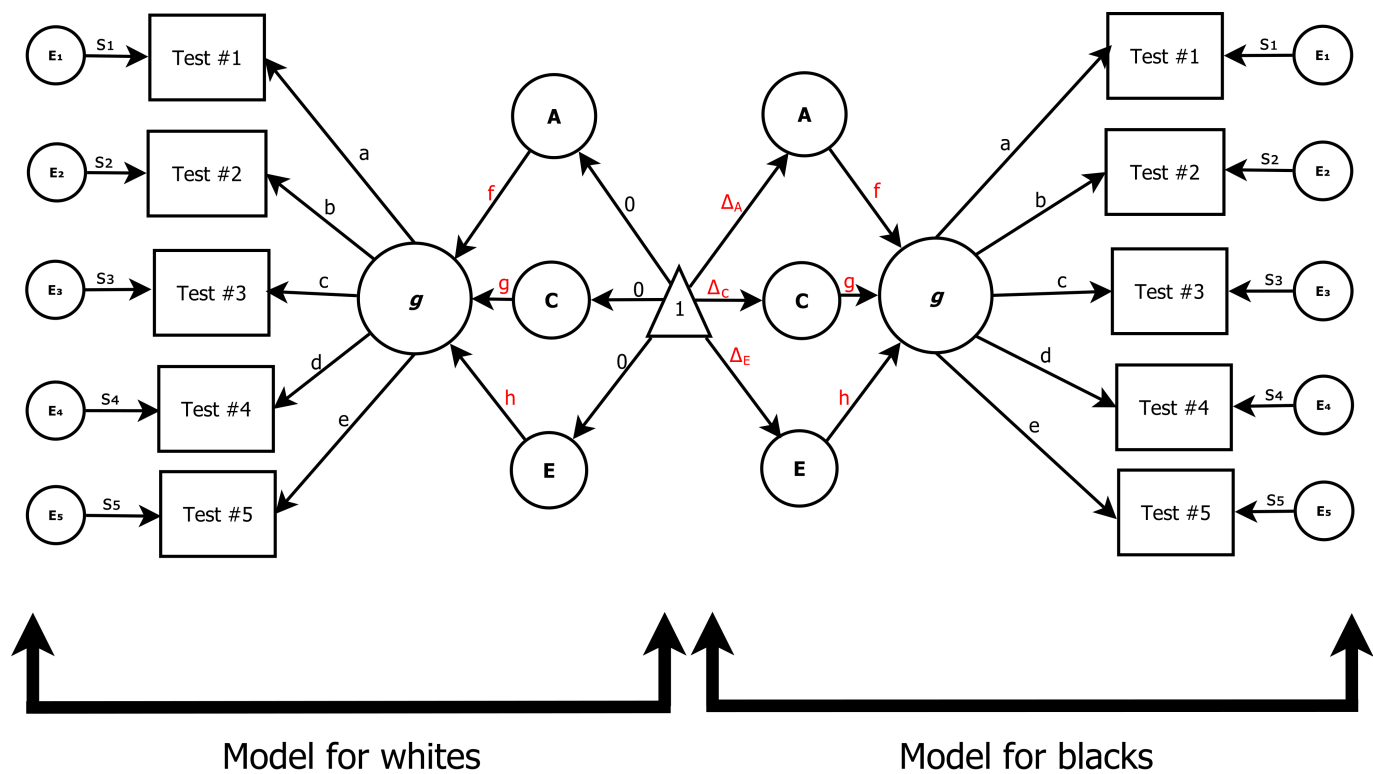
The only difference between the black and white models are the values on the paths from the triangle to the g factors. In whites, this value is fixed at 0, while in blacks this parameter, γ , is estimated in relation to the white value of 0. Blacks and whites are therefore allowed to differ in their mean value on g , and the mean difference is estimated as a contrast between the races, rather than as a difference on some absolute scale. If all black-white differences in tests #1–#5 can be explained by differences in the mean value of g (while constraining factor loadings to equality

between races, i.e., metric invariance), scalar invariance holds. Finally, it can be seen from the graph that the loadings of the tests on their unique residuals (s_1-s_5) must be constrained to equality between blacks and whites. Given that the variances of the residuals are unity, this guarantees that strict measurement invariance holds. It should be emphasized that the means of the residuals of all tests are zero. Only mean differences in g are permitted to explain mean differences in test scores between races. (The model can also accommodate race differences in the *variances* of latent abilities, but that possibility is ignored here.)

If the first stage of the model fits the data well, we can conclude that all individuals who have the same amount of latent ability have the same scores, on average, on all five tests, regardless of their racial identity. Therefore, the tests are unbiased and all black-white differences in them can be attributed to g .

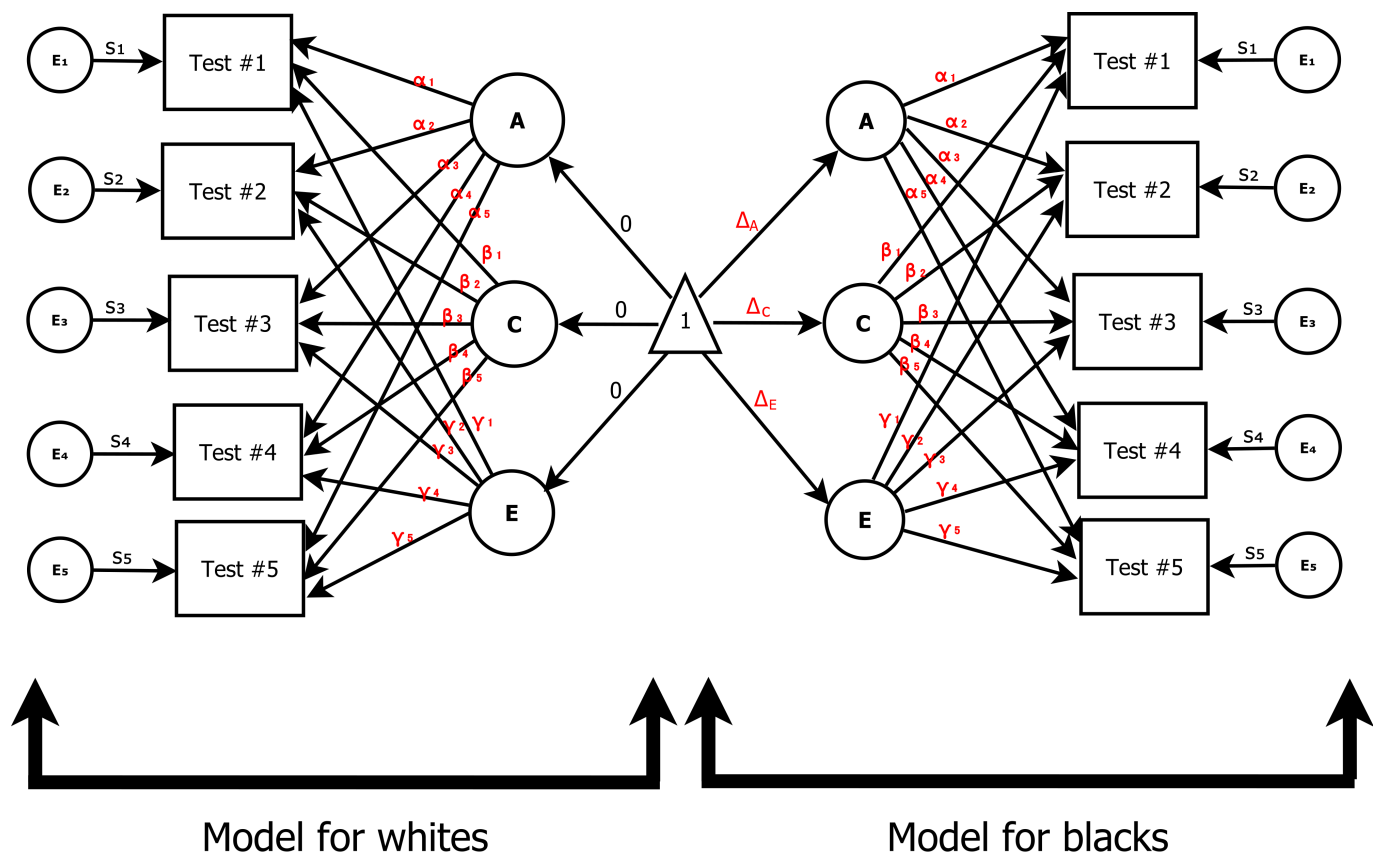
While the end point of measurement invariance investigations is normally the discovery of greater or lesser invariance and perhaps the estimation of latent mean differences, the current model would go further by examining *why* the two races differ on g . In particular, biometric modeling enables the causal attribution of mean differences to genetic causes, shared (familial) environmental causes, and other, non-shared causes (accidents and other unique experiences)—other potential sources of influence, such as gene-environment correlations and interactions could possibly also be included but they are ignored here. The proportions of the IQ gap that come from biometric sources could therefore be estimated.

The second, biometric stage of the model can be depicted as this graph (click for a larger image):



The invariant parameter values (a – e and s_1 – s_5) that were estimated in the first stage are reused in the second stage, i.e., the parameters are fixed to previously obtained values. Some other parameters are fixed to zero. The parameter values marked in red are estimated in the second stage. The biometric variables included are A, C, and E, which correspond to genetic effects, shared environmental effects, and residual effects, respectively. The graph does not specify how the biometric components are estimated, but the classical twin model would be an obvious choice.

However, the model depicted above would not in practice be identified. To estimate the contribution of biometric components to group differences, there must be at least one more phenotypic variable than the number of biometric components included. A solution to this problem would be to remove the g factor from the model and estimate genetic and environmental influences on the five tests, as in this graph:



The first stage established that any differences between whites and blacks must be due to g . The three biometric variables explain 100 percent of the g variance, and, considered together, are perfectly isomorphic with g . The values on the paths from A, C, and E to tests #1–#5 (i.e., γ_1 – γ_5 , γ_1 – γ_5 , and γ_1 – γ_5) are partial regression weights. In a standardized model, the squared values of these regression weights are equivalent to the proportions that genetic, shared environmental, and residual effects on g explain of the test's variance. For example, if the estimate for the value of the path γ_1 is .80, the proportion of variance that genetic effects on g explain of the test's variance is $.80^2 = .64$. This is not necessarily equal to the full heritability of the test because its residual can also be genetically influenced. However, the first stage of the model guarantees that there are no black-white mean differences in the residuals of any of the tests, so whatever genetic or environmental effects there are on the residuals can be ignored.

The biometric model would in practice be estimated in two steps. At first, racial mean differences in the tests are ignored, and the values of the biometric parameters for each test are estimated. It should be noted that the model requires that the sizes of the biometric parameters be the same in blacks and whites (or at least that the ratios of the unstandardized biometric variances are the same across races; see Dolan et al., 1992). This may seem like a highly limiting constraint, but it should be noted that the heritabilities and environmentalities of IQ are generally quite comparable between whites and blacks (Pesta et al., 2020). Furthermore, the determinants of a latent, race-invariant g factor would be expected to be particularly similar across races. In any case, this is something that can be tested.

After the best values for genetic, shared environmental, and residual paths for each test have been obtained through model comparison (e.g., ACE vs. AE vs. ADE models), the means are added to the model and the model is fitted again. The triangle connected to the biometric components in the graph specifies that whites and blacks may differ in the mean values of any of the three biometric components influencing g ; to facilitate model identification, white means are fixed to 0, and black means are estimated in relation to the white means. This is the crucial test of whether black-white differences in the tests and thus in g can be reproduced from means differences in the latent biometric variables. If the fit of the model does not deteriorate when means are added to it, the genetic and environmental factors that account for individual differences also account for black-white differences.

The model will give estimates of mean black-white differences on the three biometric factors. Thus, it is, in principle, capable of providing exact answers to questions at the center of the race and IQ controversy. However, it appears to be the case that if a latent, reflective g does underlie the black-white IQ gap, then the biometric part of the model delineated will be empirically unidentified. This is because the genetic and environmental loadings of the tests on the biometric factors will not provide non-redundant information; given that g is the only source of common variation in the tests, the loadings of the different tests will be the same up to a multiplicative constant.

There are several other preconditions that must be met for the model to be usable. First, the tests must have a joint normal distribution, at least if maximum likelihood with its more moderate sample size requirements is used for parameter estimation. Second, as already mentioned, there must be at least one more test available than the number of biometric components included—i.e., if A, C, and E are estimated, there must be at least four tests (however, as pointed out in Dolan et al., 1992, C and E can be collapsed into a single environmental variable that is correlated between twins, reducing the number of tests needed by one in the classical twin model). Third and perhaps more problematically, the tests must be congeneric, i.e., a single common factor must account for all of their covariances in both blacks and whites.

The demands that this model puts on the data greatly limit its usability. However, given the flexibility of the SEM framework, it seems likely that there are ways to relax some of these requirements. For example, it might be possible to bypass the congenericity requirement by using a bifactor model. This would involve specifying an invariant bifactor model of black-white differences in the first stage, as in Frisby & Beaujean (2015). In the second stage, the loadings on the non-*g* factors and factor means from the first stage would be retained, while the *g* factor would be replaced with biometric factors. It would seem that in this setting it might be possible to biometrically decompose mean differences in a *g* factor that is derived from non-congeneric indicators.

However, regardless of the congenericity requirement, it appears that the nature of *g* would make the biometric model of the means empirically unidentified, as discussed above. In technical terms, this is because the *g* model is a common pathway model while the biometric mean model is an independent pathway model (for definitions, see Franic et al., 2013). Would it nevertheless be possible to make the model identified by adding some kinds of information to it, in the way same way that certain identifying assumptions of the classical twin model can be relaxed if additional information is available (cf., Derks et al., 2006 and Dolan et al., 2019)? This is one interesting problem for hereditarians to pursue. Meanwhile, however, other approaches, such as the admixture method discussed later, will have to be used to estimate the contribution of genetic differences to the black-white gap.

2.3. Possible counterarguments to the default model

2.3.1. Race as a unit of analysis

It is often suggested that because races are not “natural kinds”, using them as units of genetic analysis is mistaken. After all, one could find reasonable justifications for lumping or splitting the genetic diversity of humanity differently from how it is done in, say, American discussions of race. The answer to this argument is, firstly, that biological taxonomy is not about “natural kinds.” Jerry Coyne [pointed out](#) that it is easy to justify human racial differentiation if one uses the usual biological criteria, rather than the non-biological criteria of “naturalness”:

What are races?

In my own field of evolutionary biology, races of animals (also called “subspecies” or “ecotypes”) are morphologically distinguishable populations that live in allopatry (i.e. are geographically separated). There is no firm criterion on how much morphological difference it takes to delimit a race. Races of mice, for example, are described solely on the basis of difference in coat color, which could involve only one or two genes.

Under that criterion, are there human races?

Yes. As we all know, there are morphologically different groups of people who live in different areas, though those differences are blurring due to recent innovations in transportation that have led to more admixture between human groups.

How many human races are there?

That’s pretty much unanswerable, because human variation is nested in groups, for their ancestry, which is based on evolutionary differences, is nested in groups. So, for example, one could delimit “Caucasians” as a race, but within that group there are genetically different and morphologically different subgroups, including Finns, southern Europeans, Bedouins, and the like. The number of human races delimited by biologists has ranged from three to over thirty.

I like the pragmatic perspective advocated in Fuerst (2015), according to which a biologically valid unit of analysis is one that is deeply enmeshed in some research program. If you want to argue that race is not a valid variable to use in such research, abstract conceptual arguments have no bite; you have to actually show that the actual way race is used in a research program is invalid. In hereditarian research on the black-white IQ gap the only thing that really needs to be true about race is that self-identified whites differ in terms of many allele frequencies from self-identified blacks—and this is something that is trivial to demonstrate these days (graph from Kirkegaard, 2019):

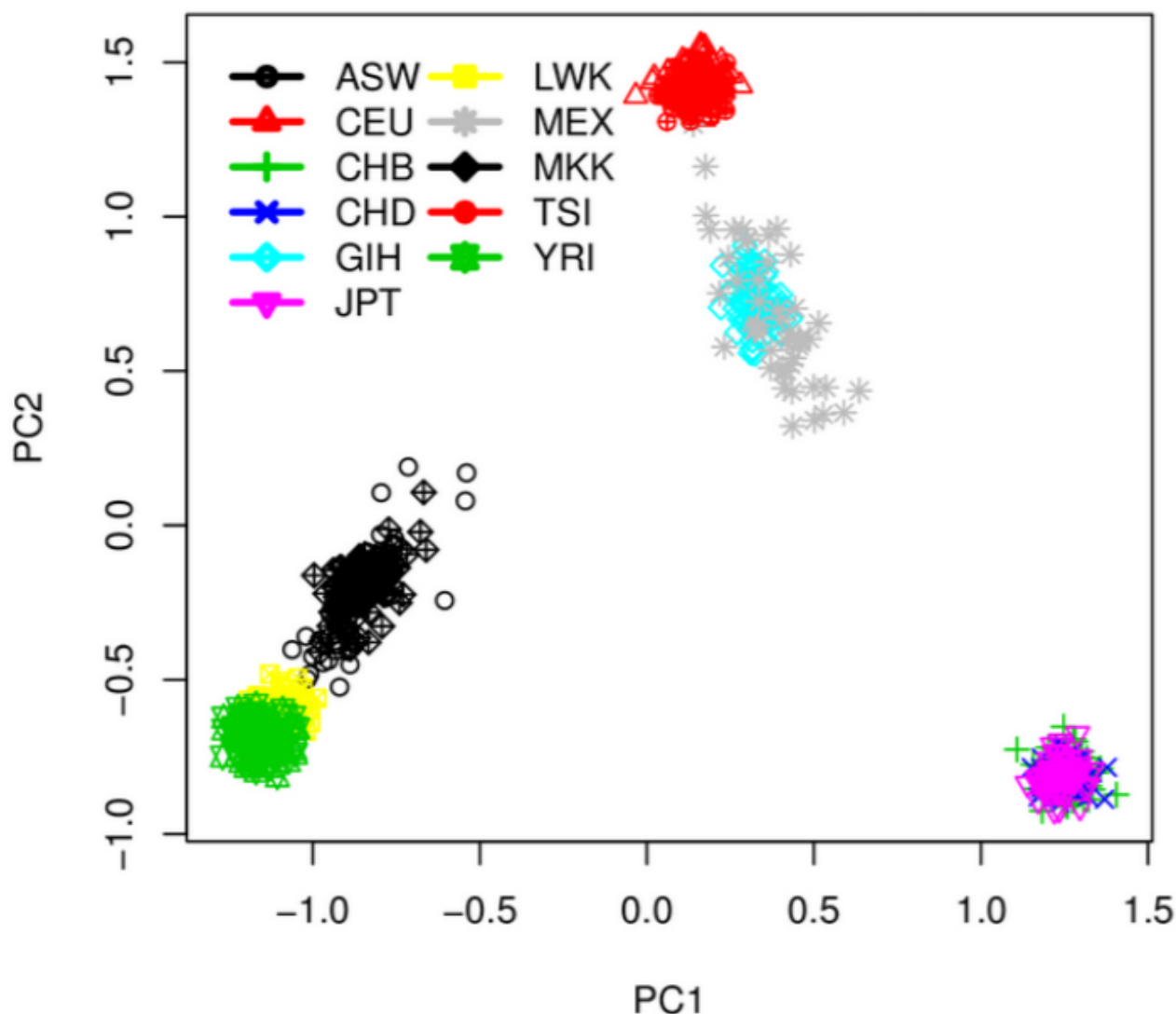


Figure 1. Principal components analysis of HapMap3 data. Each dot is a person. Axes are the first two principal components of the SNP (single nucleotide polymorphism) data. Populations: ASW = African American in USA; CEU = Central European from Utah, USA; CHB = Chinese from Beijing; CHD = Chinese from Denver, USA; GIH = Gujarati Indians in Houston, USA; JPT = Japanese from Tokyo, Japan; LWK = Luhya Africans from Kenya; MEX = Mexicans from Los Angeles, USA; MKK = Maasai Africans from Kenya; TSI = Italians from central Italy (Toscani); YRI = Yuruba Africans from Nigeria. Figure reproduced from Abraham and Inouye (2014).

Given that “everything is heritable” (Polderman et al., 2015), almost any socially salient group will differ from others due to inherited causes in at least some ways. That is why I have never found the “race is not real” argument even remotely persuasive. Much less biologically “real” categories than race are perfectly suitable for genetic analysis.

Another pragmatic reason for using race in genetic analyses is that race is one of the central explanatory variables used by non-hereditarians in American social science. You cannot foreground race in the analysis of all social problems and expect others to completely ignore its obvious genetic correlates.

2.3.2. Are hereditarians overtly egalitarian?

A possible, rather ironical counterargument to the default model follows from the fact that many genetic variants causally associated with IQ are rare and may be “private” to specific populations. Whereas the hereditarian model posits that race differences in intelligence are quantitative (all races have the same abilities, with substantial differences only in central tendencies and, possibly, variances), differences in the genetic architecture of cognition open up the possibility of deeper, essential differences. For example, the analysis by Hill et al. (2018) suggests that close to half of the genetic contribution to IQ variation is due to single nucleotide polymorphisms with a minor allele frequency in the range of 0.001–0.01. Many or most variants that are so rare may be private to specific populations. If the genetic basis of intelligence differences is substantially different in different races, is it reasonable to assume abilities to be at all comparable between races? Have hereditarians underestimated the nature and extent of racial differences?

I think that the challenge of this kind of causal heterogeneity is more apparent than real. Genetic differences between races are of interest to behavioral science because of their downstream, phenotypic effects—on cognitive ability in this case. Downstream effects are a *psychological* and *psychometric* rather than genetic issue, and we already know that, to the best of our knowledge, cognitive ability does not differ qualitatively between whites and blacks. The fact that a certain causal genetic locus is polymorphic in one population while the same locus is homozygous in all other populations does not threaten the validity of population comparisons for the phenotype in question. From a causal perspective, such loci are no different from loci that are fixed versus polymorphic among different subpopulations of the same population. For example, if the average additive effect of some allele on IQ is x versus 0 for the other allele, then the effect of the (biallelic) locus in members of the “polymorphic” population is either 0, x , or $2x$ while in all members of other, “fixed” populations the effect is 0 or $2x$, depending on which allele is fixed.

I think that causal heterogeneity is true at least to some extent, but that it does not pose a conceptual problem for the default model. It does, however, make genetic variant discovery harder. The ability to explain all heritable group differences in terms of individual alleles is in the distant future. Therefore, methods such as twin studies and admixture analyses that can indirectly estimate the total effect of all genetic variants remain important.

2.3.4. Modifiability of highly heritable characteristics

It is often said that high heritability does not imply that the phenotype is not modifiable. This is true in a purely logical sense: genes work through physical mechanisms in which we can in principle intervene just like in any other physical mechanisms. However, something being possible in principle is quite different from it being actually possible here and now. We know that a specific dietary intervention helps allay the negative cognitive effects of [phenylketonuria](#), but that is only the case because we have some mechanistic understanding of this rare Mendelian condition. In contrast, we certainly do not understand how the thousands of genetic polymorphisms that are involved in the normal, quantitative variation in IQ exert their influence. Understanding and changing the functioning of up to thousands of polymorphisms, each with a tiny effect on the phenotype, is a very different undertaking than manipulating a single large-effect Mendelian trait—and, in fact, most Mendelian disorders are much less intelligible and treatable than phenylketonuria.

In practice, most attempts to boost intelligence are educational and cognitive-psychological in nature. They do not rely on any molecular-level understanding about the mechanisms of intelligence. Instead, they are based on various hunches about how intelligence might possibly work. The track record of these efforts in producing permanent ability gains is predictably poor (e.g., Protzko, 2015; Sala & Gobet, 2017). This has not stopped people from discounting heritability as uninteresting or as lacking in policy relevance. In effect, they have argued that it is sufficient that we can *mentally conceive* of environmental changes that would equalize outcomes. Sesardic (2005, p. 84–85) noted that such arguments represent the “curious triumph of the *possible* over the *actual*.” I think we should dismiss the modifiable-in-principle argument as nonsense, just like we dismiss people promising miracle cures to currently incurable diseases.

Jensen suggested in his 1969 paper that IQ enhancement is more realistically in the purview of biology rather than psychology or education. From a hereditarian perspective, it is certainly possible to equalize white and black IQ distributions if one can resort to eugenics. There is plenty of IQ-linked genetic variation in blacks, and instituting policies that bias black reproduction so that high-IQ parents will have more children than low-IQ parents will inevitably lead to the closing of the black-white gap over time, assuming that white fertility remains less eugenic. In effect, this is what has happened in the case of some recently established black diaspora populations in Western countries. Because of immigrant selection based on educational credentials and financial resources, some expatriate populations (and their offspring) are substantially smarter, on average, than their coethnics in the old country.[\[Note\]](#)

2.3.5. Flynn effect: The revolution that fizzled out

In the last few decades, environmental or nurturist views have dominated social science, but this apparent victory of nurture over nature has much more often been based on skillful rhetoric rather than any firm empirical findings or theoretical breakthroughs. In contrast, hereditarianism, while

constantly facing strenuous (but mainly rhetorical) challenges, has gone from strength to strength as an empirical research program, especially when it comes to the individual differences paradigm of mainstream behavior genetics.

Nurturism has nevertheless achieved one apparently great triumph: the discovery that raw scores on IQ tests have rapidly increased over the last 100 years in numerous countries, with average gains between oldest and youngest cohorts amounting to dozens of standard score points in many tests. The increases are so widespread that they appear to be part and parcel of the general process of modernization. On the face of it, the Flynn effect, as the phenomenon is called, appears to affirm the wildest nurturist optimism, showing the tremendous power of environmental improvements to boost intelligence. By the same token, it appears to challenge the validity of research that stresses the high heritability and stability of IQ. The increases are too large for heredity to have much to do with them. Given that the raw score gap between, say, grandparents and grandchildren in many tests is as large as the black-white gap, the Flynn effect would also seem to suggest that the racial gap can be closed through a suitable mix of environmental remedies.

However, a closer look at the score gains indicates that the Flynn effect cannot be interpreted as a straightforward increase in intelligence. While some of the gains undoubtedly reflect genuine, biological improvements, paralleling simultaneous increases in average height, most of the gains have an artefactual quality. In particular, when methods originally developed to find tests and items that are biased against cultural or racial minorities are applied to test results from different cohorts, it becomes abundantly clear that IQ tests are typically biased against members of older cohorts.

My view is that IQ tests measure strongly genetically influenced individual differences in the capacity to absorb information available in a given broad cultural environment and to apply that information in different contexts. If cultural environments differ widely between groups, a given IQ test may not be an unbiased measure of those underlying individual differences. Importantly, it is possible to find out if environments differ between groups so greatly that IQ tests cease to tap into the same capacities. This is possible in the context of latent variable models. In particular, an investigation of measurement invariance will reveal whether an IQ test provides comparable estimates of the abilities of different groups.

My interpretation of *strict* measurement invariance is that if it holds between two groups for IQ, then the groups must have been exposed to very similar environments and have had opportunity to absorb the same knowledge and skills. Differences in IQ between groups must then largely reflect differences in underlying, strongly genetically conditioned cognitive capacities. Once we have ascertained that groups are being compared on an invariant scale, classic findings about IQ—regarding group differences, predictive validity, heritability, stability, and so on—can be fully affirmed.

It is well-established that scalar and strict measurement invariance are generally untenable in between-cohort IQ comparisons (Wicherts et al., 2004; Beaujean & Osterlind, 2008; Must et al., 2009; Wai & Putallaz, 2011; Shiu et al., 2013; Pietschnig et al., 2013; Fox & Mitchum, 2013, 2014; Beaujean & Sheng, 2014). On the other hand, black-white IQ differences in America generally meet the requirements of strict invariance (e.g., Dolan, 2000; Dolan & Hamaker, 2001; Lubke et al., 2003; Trundt, 2013; Frisby & Beaujean, 2015; Lasker et al., 2019). It has also been found that Flynn gains are of a very similar magnitude in blacks and whites (Ang et al., 2010), indicating that whatever environmental changes are involved in the gains, whites and blacks are not differentially exposed to them (at least for the cohorts born in the 1980s and 1990s studied by Ang et al.). In short, IQ differences between blacks and whites (of the same cohort) can be treated as genuine ability differences, whereas the same inference is not justified when comparing different cohorts of people.

The lack of measurement invariance does not necessarily imply that no real ability gains are involved in the Flynn effect. It simply means that if at least a subset of available tests are not invariant, IQ differences between groups cannot be interpreted in the way you would interpret differences within a group. It is surely the case that the knowledge and skill level of the population has improved at least in some domains as the time that the average person spends in school has increased by up to hundreds of percents. Nevertheless, interpreting IQ gains as gains in *intelligence* is problematic. Even aside from technical psychometric concerns about invariance, this is apparent when we simply consider the tails of the IQ distribution. If the true score distribution of IQ had really shifted by dozens of points or more in a few generations, a quarter or more of the population within living memory should have had incapacitating mental disabilities preventing independent living. But that was not the case. Looking at the other end of the IQ distribution, if intelligence levels had really skyrocketed, the greatest intellectual and scientific achievements of a hundred years ago could today be quite easily understood by the average person. Yet it is far from true that the physics of an Einstein or the mathematics of a Hilbert can now be easily grasped by the average person.

James Flynn has amply earned his fame for the diligent work he carried out to document the universality of his eponymous effect. Disturbed by the implications of Jensen's research program, he faced head-on the challenges that it posed to conventional social science. Unlike most critics of hereditarianism, he did not dismiss any hypotheses on political or ideological grounds, nor did he try to come up with clever conceptual arguments so as to ward off the whole debate. Instead, he collated massive amounts of novel data that called into question core hereditarian beliefs. Nevertheless, this turned out to be much less of a revolution than it initially appeared to be. The lesson to learn from the Flynn effect is ultimately not about the malleability of intelligence but rather about the challenges of accurately measuring the abilities of culturally divergent groups, including different cohorts of people within a single, rapidly changing society.

2.3.6. Realism run amok?

One more counterargument to the model of the black-white gap that I have described could be that it relies on realist interpretations of various complex theoretical constructs whose ontological statuses have been disputed. For example, while the g factor is widely accepted as a causal entity by intelligence experts, this acceptance is not universal. However, I make no apologies about this aspect of the model. My perspective is that boldness in theoretical formulations is a virtue, enabling strong empirical tests.

Furthermore, if one rejects the g factor model for the black-white gap, there remains the issue of why the differences track g differences, and why the quite demanding conditions of measurement invariance typically hold between these two races in America. Why are between-race IQ differences statistically indistinguishable from within-race IQ differences if they arise from different causal processes? It is not enough to deny the truth of the default model. One must show that another model does a better job at explaining the facts at hand.

3. Bradford Hill criteria and the black-white gap

3.1. Inferring causes

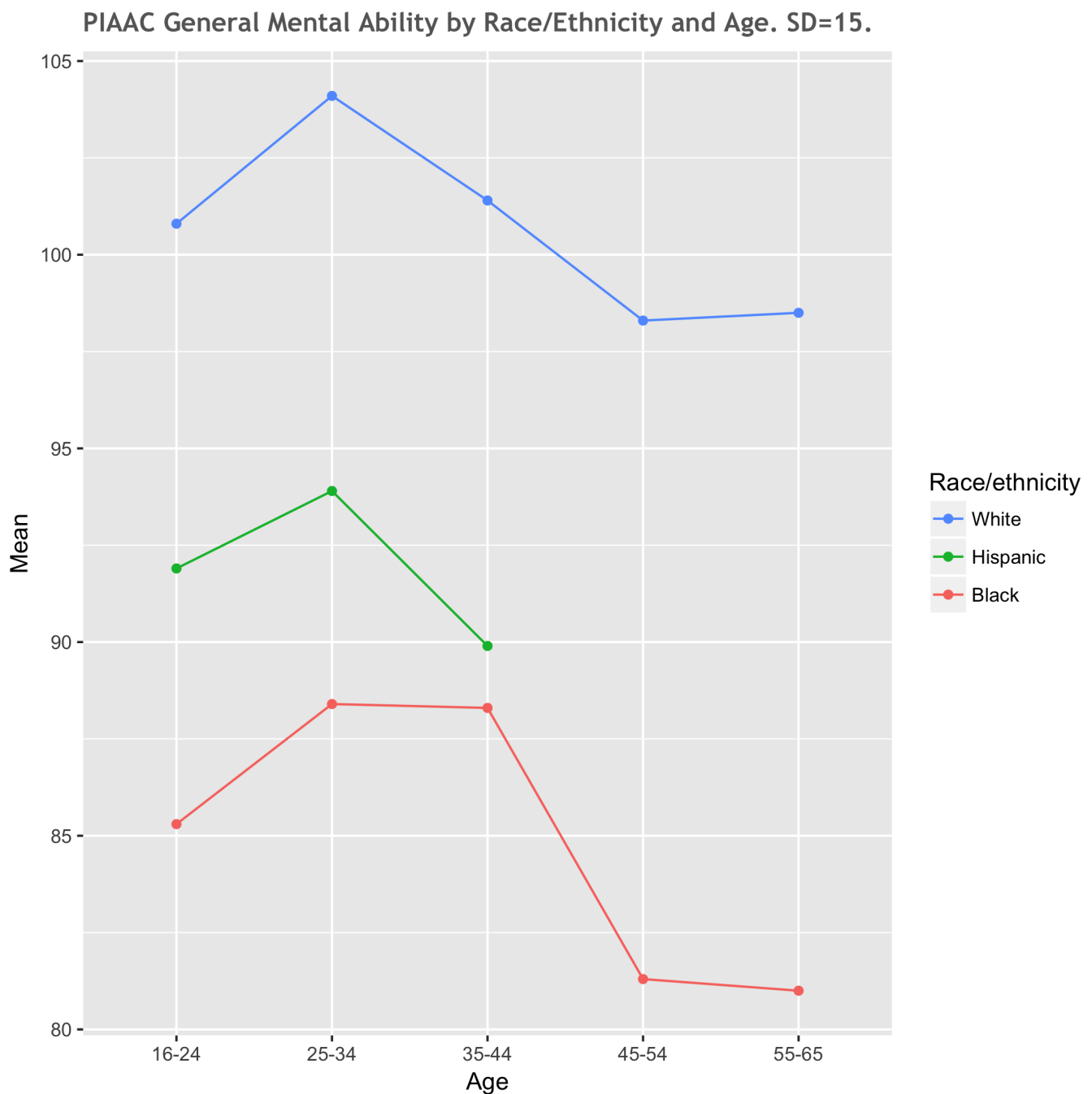
In a 1965 article, the English epidemiologist Austin Bradford Hill promulgated a set of nine principles that can be used to help establish causality in areas of research where most of the available evidence is correlational rather than experimental in nature (Hill, 1965). He had in mind diseases and their environmental causes, such as lung cancer and smoking, but I think most of his criteria are quite apposite for evaluating the hereditarian thesis of black-white differences in IQ, too. Below, I use Wikipedia's [summaries](#) of the nine criteria to elaborate on how each criterion relates to the hereditarian thesis.

3.1.1. Strength (effect size)

A small association does not mean that there is not a causal effect, though the larger the association, the more likely that it is causal.

Research over the last 100 years indicates that the black-white IQ gap in the US is about one standard deviation, or 15 points. This effect size was first found in military IQ tests administered during the First World War. Perhaps the first truly systematic investigation of the question was that of Audrey M. Shuey whose book *The Testing of Negro Intelligence* (1958; 2nd ed., 1966) concluded that a one standard deviation black-white gap was found from the 1910s through the 1960s. Jensen's investigations confirmed this effect size as did a 2001 meta-analysis with a sample size of several million people (Roth et al., 2001). It has frequently been claimed that the gap is decreasing (e.g., Dickens & Flynn, 2006), but more searching analyses fail to confirm the claim (Murray, 2007). It seems plausible that the gap was *larger* than one standard deviation at least in some places in the past, but there is scant evidence for a narrowing of the gap to a level below one standard deviation. I analyzed the nationally representative PIAAC sample, which was

recruited in 2011–2014, [here](#), and found that the gap was around one standard deviation regardless of generation:



An effect size of one standard deviation is large by most standards. In a recent survey article on effect sizes in psychological research, Funder and Ozer (2019) argued that a Pearson correlation of .40 or greater is “a very large effect”, and claimed that an effect of that magnitude is “likely to be a gross overestimate that will rarely be found in a large sample or in a replication.” A correlation of .40 is equivalent to a standardized group difference of $d = 2 * .40 / \sqrt{1 - .40^2}$? **.87**, so the black-white gap is clearly *a very large effect* by this standard. Nevertheless, it is an effect that has been repeatedly found in large samples for a century now.

In practical terms, the $d = 1$ gap is similar in magnitude and largely coterminous with the educational achievement gap between whites and blacks as seen in, for example, [NAEP tests](#) and the [SAT](#). The $d = 1$ gap is also large enough an effect to statistically [account for the wage gap between whites and blacks](#). It can also largely explain differences in intergenerational economic mobility between blacks and whites (Mazumder, 2014). Given that IQ has the most extensive domain of predictive validity of all variables in social science, there are few commonly discussed black-white gaps where it is not involved.

The mere sizableness of the black-white gap does not, of course, show that it must be genetic in origin. However, the large effect size poses problems for the hypothesis that the gap is a mere contingency of environmental conditions. According to the hereditarian model, the IQ gap is mostly inherent to blacks and whites, and its existence is not dependent on people's particular circumstances. In contrast, environmental explanations must posit that there is a high correlation between race and environmental exposures of some kind. The simplest environmental model is that all blacks and no whites are victims of racial discrimination that causes a loss of fifteen IQ points. However, as pointed out by Flynn (1980, p. 60), this model does not make conceptual or empirical sense:

Racism is not some magic force that operates without a chain of causality. Racism harms people because of its effects and when we list those effects, lack of confidence, low self-image, emasculation of the male, the welfare mother home, poverty, it seems absurd to claim that any one of them does not vary significantly within both black and white America.

More realistic models have to postulate that the relevant environmental exposures are not uniformly present or absent in blacks or whites, but rather that they are operative to a greater extent in blacks and to a lesser extent in whites (or vice versa for positive influences). This means that racial differences in environments must be substantially *greater* than one standard deviation in magnitude to explain the full IQ gap. In practice, however, it is not easy to statistically control the gap away even in large regression models with many independent variables, let alone to show that IQ is causally downstream of such variables. Environmentalists must simultaneously maintain that the black-white gap is caused by some of the most potent forces ever encountered in the social and behavioral sciences, and that those forces are largely unknown and harder to detect than just about anything else. Given the size of the gap, environmentalists face a similar explanatory challenge as those who denied a causal link between smoking and lung cancer. Hill (1965) put it this way:

But to explain the pronounced excess of cancer of the lung in any other environmental terms requires some feature of life so intimately linked with cigarette smoking and with

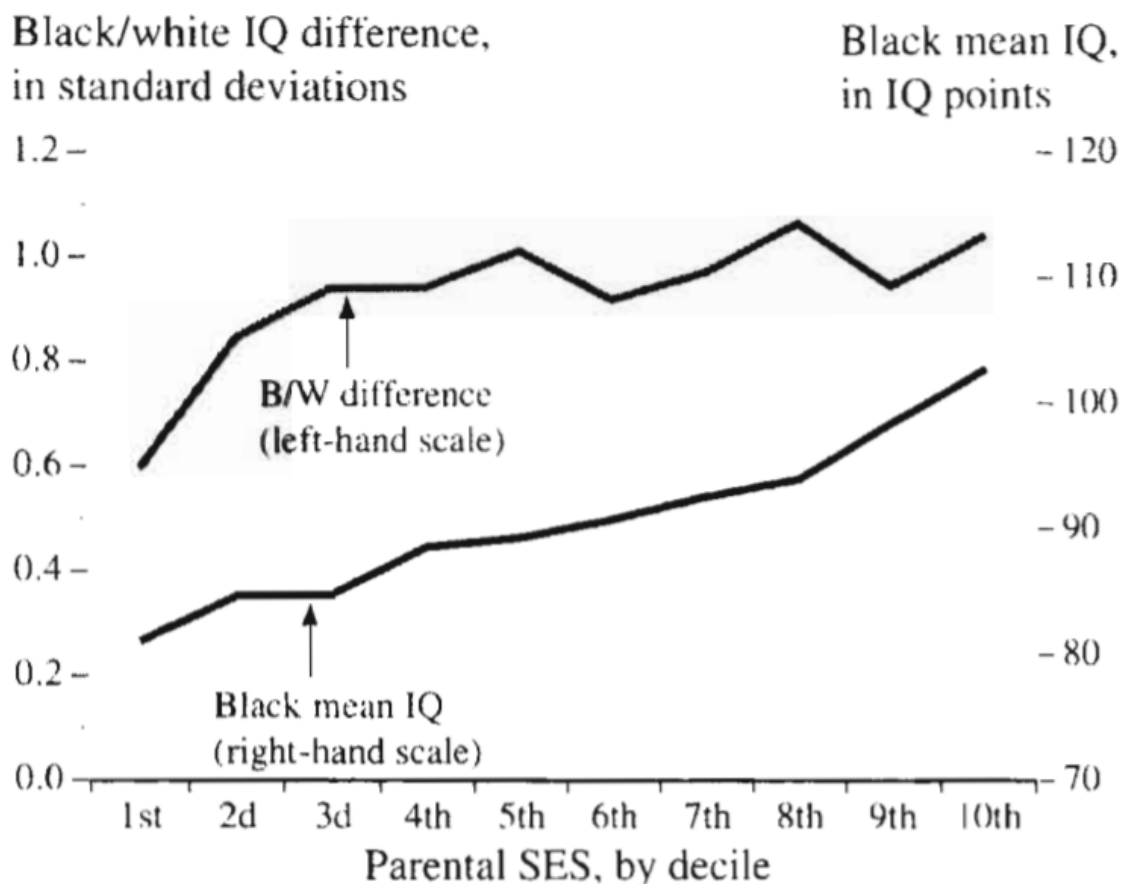
the amount of smoking that such a feature should be easily detectable. If we cannot detect it or reasonably infer a specific one, then in such circumstances I think we are reasonably entitled to reject the vague contention of the armchair critic ‘you can’t prove it, there may be such a feature’.

3.1.2. Consistency (reproducibility)

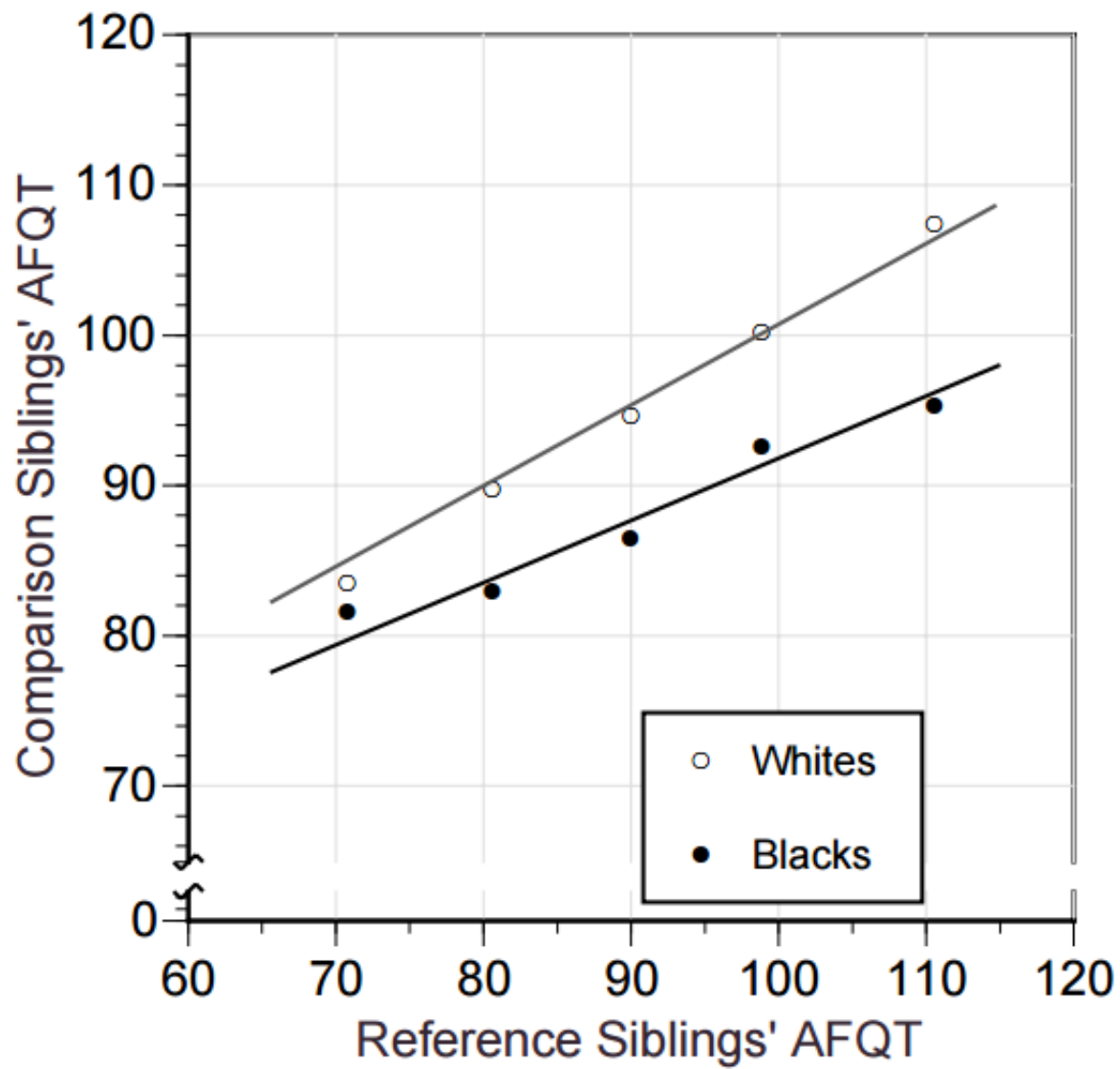
Consistent findings observed by different persons in different places with different samples strengthens the likelihood of an effect.

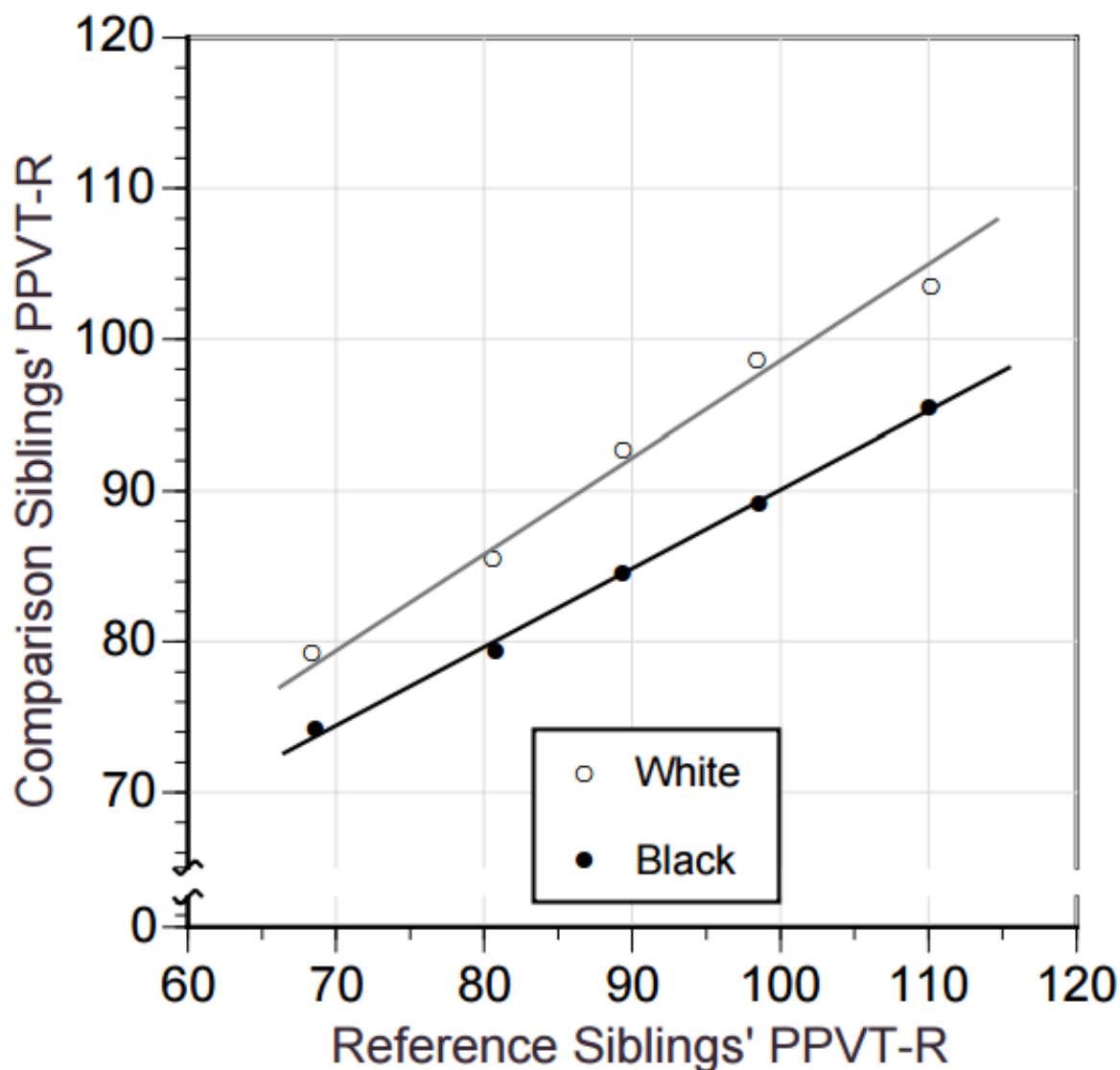
Aside from the temporal consistency, the black-white gap is persistent along other dimensions as well. For example, the gap is observed across social classes (from Herrnstein & Murray, 1994, p. 288):

Black IQ scores go up with socioeconomic status, but the black-white difference does not shrink

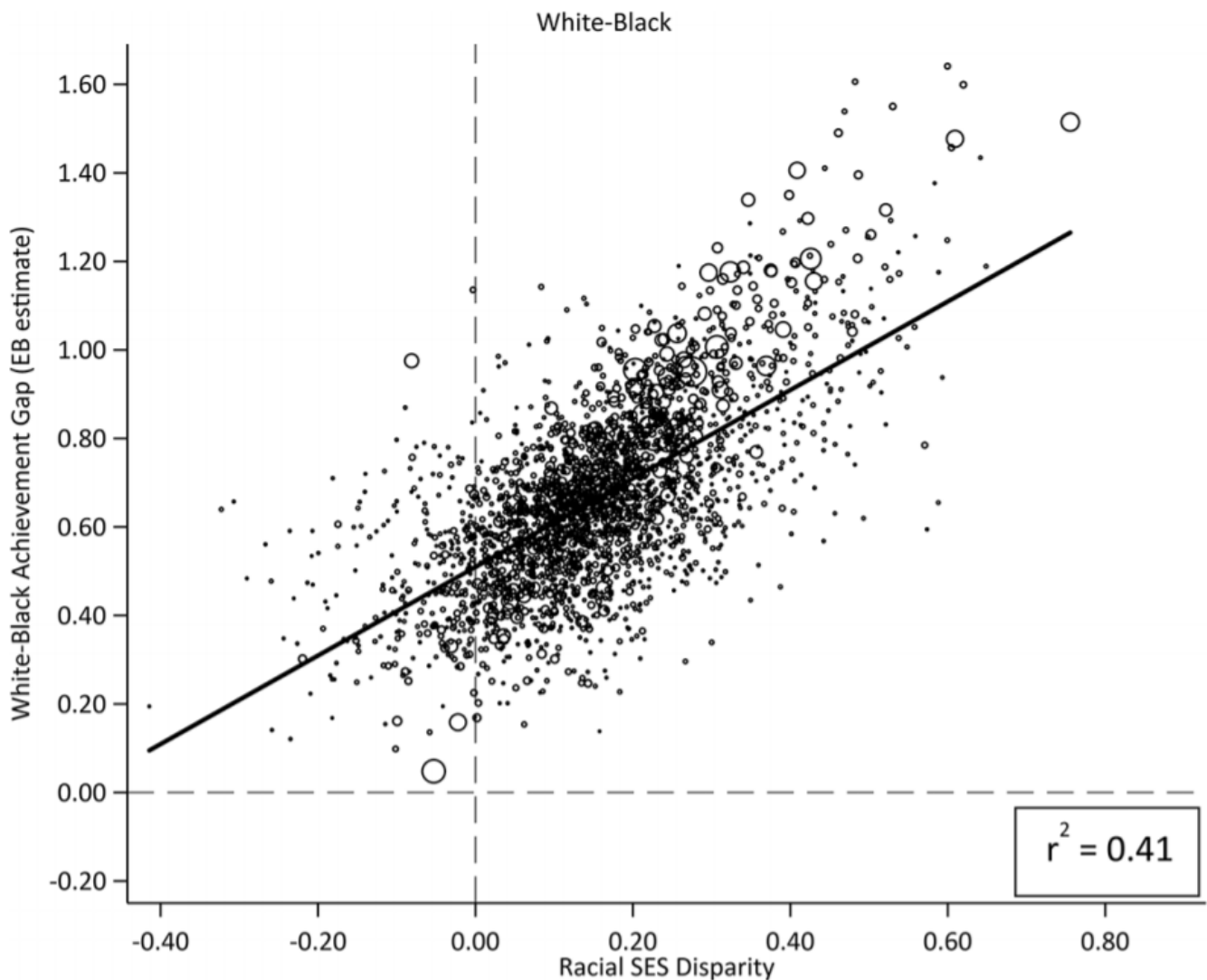


We can also predict the magnitude of the black-white gap from within-family sibling differences in IQ. This accords well with the regression effect that is expected due to non-perfect heritability and due to siblings sharing only about 50 percent of their genetic variation for a given trait. While *ad hoc* environmental models that account for this effect can be devised, I have never seen one formalized. Here are two examples of the regression effect from Murray (1999), with each dot representing between a few dozen to a few hundred sibling pairs:





The most thorough analysis of the geographical generalizability of test score gaps between whites and blacks in America is that of Reardon et al. (2019). They looked at the gaps in standardized math and English language arts (ELA) tests administered to public school students in several hundred metropolitan areas and several thousand school districts from 2009 to 2013. The analysis included results from some 200 million tests. While the tests were not IQ tests proper, standardized academic tests are good proxies for IQ (Deary et al., 2007; Kaufman et al., 2012). The following graph summarizes their findings:



About 2,300 school districts had sufficient numbers of white and black students to permit gaps to be estimated. Every single district had a gap favoring white students. The graph sorts the gaps according to racial disparity in parental SES, and we can see that while the SES disparity moderates the gap (as predicted by the hereditarian model^{[[Note](#)]}), the white advantage exists also in

all the districts where black parents have higher SES (in accordance with the regression effect that is an elemental part of the hereditarian model^[Note]). Given that no school district, no matter how selected its population or how progressive its policies, has managed to eliminate the gap, the goal of equalizing outcomes across the nation clearly has a fanciful, utopian character. The universality of the gaps across very heterogeneous environments suggests that highly [canalized](#) genetic differences are plausibly involved.

An interesting fact is that the black-white gap tends to be close to one standard deviation even in many contexts where the individuals being compared have been cognitively selected. For example, in 2005–2008 the gaps in various graduate school admissions tests (which are IQ tests, albeit ones biased toward specific item contents) varied between .92 and 1.28 standard deviations:^[Note]

Test	Prospective degree	Black-white gap (<i>d</i>)
GMAT	M.B.A.	1.13
GRE-Verbal	Ph.D./M.A.	0.92
GRE-Quant	Ph.D./M.A.	1.08
LSAT	J.D.	1.17
MCAT-Verbal	M.D.	1.28
MCAT-Phys Sci	M.D.	1.08
MCAT-Biol Sci	M.D.	1.28
DAT	D.D.S.	0.99

Yet another way to test to the generality of the black-white gap is to look at the tails of the distribution. For example, we can look at the [top scorers on the Law School Admission Test](#):

In 2004, 10,370 blacks took the LSAT examination. Only 29 blacks, or 0.3 percent of all LSAT test takers, scored 170 or above. In contrast, more than 1,900 white test takers scored 170 or above on the LSAT. They made up 3.1 percent of all white test takers. Thus whites were more than 10 times as likely as blacks to score 170 or above on the LSAT. There were 66 times as many whites as blacks who scored 170 or above on the test.

We can also examine data from a large sample of students who took the SAT in 2001, as tabulated in [this Educational Testing Service report](#):

Number of Students in Each Selected SAT Score Range

Group	SAT (V + M) Score Range		
	High (≥ 1300)	Middle (1000-1050)	Low (< 760)
African American	1,561	9,195	36,296
Asian	14,250	8,486	8,146
Hispanic	3,012	9,624	21,281
White	76,065	86,230	35,538

In 2001, a SAT score of 1300 was at the 97th percentile in the black distribution, while it was only at the 63rd percentile in the white distribution. Selective colleges that want their student bodies to “look like America” have no choice but to put a heavy thumb on the scale in favor of black applicants.

Another example is given by Giessman et al. (2013) who studied an unnamed Midwestern school district. Several whole grade cohorts completed the CogAT6 test battery for the purposes of identifying gifted students. 0.4 percent of the 1,217 black students participating scored above the 95th percentile, while 6.9 percent of the 3,665 white students did. The fact that the black-white gap is observed in the top percentiles of the IQ distribution, too, indicates that there are no privileged subgroups of blacks who are able to escape the reach of the general mechanisms that cause the gap. Racial differences in the upper tail of the distribution can be predicted well from differences in the means and variances, suggesting that a mechanism akin to polygenic inheritance is at work.[\[Note\]](#)

The persistence of the IQ gap—and, by extension, gaps in educational performance, occupational attainment, and so on—across very different contexts indicates that its causes must be ever-present. While this is naturally true of genetic differences, it tends to make environmental explanations of the gap strained. Using Nassim Nicholas Taleb’s term, you could say that the

black-white gap is [Lindy](#). The permanence of the gap in the face of all the fervent hopes to eliminate it, all the myriad policies, interventions, and programs, and the radically altered political and social realities since the civil rights movement suggests that the phenomenon is not a contingent fact of contemporary circumstances but rather something that has persisted and will persist for much longer than any environmentally-oriented social scientist or reformer has anticipated.

3.1.3. Specificity

Causation is likely if there is a very specific population at a specific site and disease with no other likely explanation. The more specific an association between a factor and an effect is, the bigger the probability of a causal relationship.

Environmental hypotheses about the causes of the black-white IQ gap suppose that blacks have been exposed to various negative environmental influences (or *not* have been exposed to positive influences) that have a large effect on the brain and behavior. A problem with this thesis is that blacks do not suffer from a generalized deficiency across different psychological functions. Rather, the deficiency compared to whites is localized to intellectual functions and specifically to the general factor of intelligence. There is no evidence of large black deficits in most psychological domains unrelated to IQ. In fact, blacks appear to be, if anything, in some ways more mentally healthy and confident about their abilities than whites. I once [collated some data](#) to demonstrate this state of affairs:

Table 1. Racial differences in emotional well-being, aspirations, and IQ.

Variable	Black-white difference (<i>d</i>)	Reference	Notes
Self-esteem	0.19	Twenge & Crocker (2002)	Meta-analysis, K=354
Positive affect	0.05	original analysis	Analysis of public-use data from Wave IV of the Add Health study; representative national sample, ages 24–34; positive affect operationalized as in De Neve & Oswald (2012); variables used: H4MH20, H4MH24, H4MH25, and GSWGT4_2 (weights)
Educational aspirations	–0.01	Mau & Bikos (2000)	Representative national sample of adolescents and young adults (males and females pooled)
Occupational aspirations	0.00	"	"
Panic disorder	0.28	Breslau, et al. (2006)	Representative national sample, odds ratio for lifetime risk converted to <i>d</i>
Generalized anxiety disorder	0.28	"	"
Social phobia	0.12	"	"
Post-traumatic stress disorder	0.00	"	"
Bipolar disorder	–0.10	"	"
Dysthymic disorder	–0.13	Riolo et al. (2005)	Representative national sample, age 15–40, unadjusted odds ratio converted to <i>d</i>
Major depressive disorder	0.27	"	"
Female body dissatisfaction	0.33	Roberts et al. (2006)	Meta-analysis of "global" body dissatisfaction, K=38
Suicidal thoughts or attempts	0.15	Blum et al. (2000)	Representative national sample of 7th to 12th graders, odds ratio converted to <i>d</i>
Suicide rate	0.43	Williams & Jackson (2005)	Death from suicide in 2000, National Center for Health Statistics data, odds ratio converted to <i>d</i>
IQ	–1.10	Roth et al. (2001)	Meta-analysis of racial differences in general cognitive ability, K=105

Note: When *d* > 0.00, blacks are better off than whites.

The relative racial equality that obtains in many non-intellectual domains is not a problem for the hereditarian model. This is because hereditarians do not claim that blacks suffer from any kind of general psychological encumbrance. Rather, the hereditarian claim, as formulated by Jensen, is that there is a gap in the general factor of intelligence that can be identified from any reasonably sized IQ test battery. Black-white similarity and gaps favoring blacks in non-*g* domains, including intellectual ones, are entirely compatible with the hereditarian model. For example, Frisby and Beaujean (2015) found the following black-white gaps on various mutually uncorrelated latent ability factors derived from some twenty Wechsler ability tests:

Table 8

Black–White mean differences on latent variables.

Factor	Estimate	SE	95% CI	
			Lower	Upper
<i>g</i>	1.16	0.10	0.97	1.34
Verbal Comprehension	0.23	0.11	0.01	0.45
Visual Processing	0.80	0.15	0.51	1.08
Working Memory	0	–	–	–
Processing Speed	0	–	–	–
Long-Term Retrieval	–0.35	0.16	–0.65	–0.04

Note. Estimates came from model M6a (see Table 6). Latent mean differences for Working Memory and Processing Speed were constrained to zero. For all latent variables, the variances were fixed at 1.0 and the means for the Black group were fixed at 0.0. Thus, a positive difference indicates the average score from the White group was higher, while a negative difference indicates the average score from the Black group was higher.

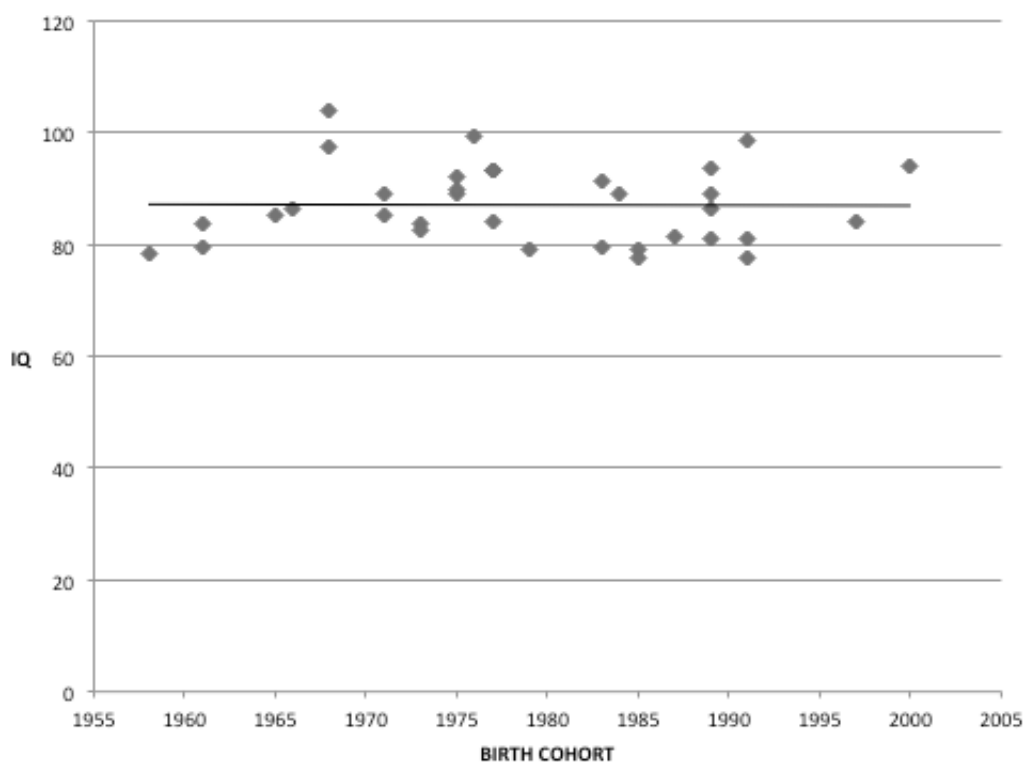
Aside from the 1.16 standard deviation *g* gap, the study found gaps favoring whites on Verbal Comprehension and Visual Processing, and a gap favoring blacks on Long-Term Retrieval. There were no significant racial differences on the Working Memory and Processing Speed factors. (Note that because the variance explained by the non-*g* factors is small, and because any real-world intellectual performance taps into both *g* and specific abilities, specific, as opposed to general, ability differences between blacks and whites are less salient in everyday life.) These results fit comfortably with the hereditarian model, while for the environmentalist the curious specificity of the black deficit in the face of the universal environmental adversity that blacks are assumed to experience is more of a mystery.

3.1.4. Temporality

The effect has to occur after the cause (and if there is an expected delay between the cause and expected effect, then the effect must occur after that delay).

According to the hereditarian model, the black-white gap originates mostly from racial differences in allele frequencies with which each new generation is born. Hill's temporality criterion is therefore not directly relevant for the model, but the criterion does suggest various tests of the model. This is because genetic and environmental effects on IQ are to some extent age-dependent. In particular, it is known that heritability increases and shared environmental influences decrease with age (Haworth et al., 2010; Tucker-Drob & Briley, 2013). The hereditarian therefore expects the gap, or at least its causal sources, to change as children grow up.

It seems that the full black-white gap appears quite early in development, perhaps as early as IQ can be feasibly measured. Jason Malloy [found](#) the full gap at age 3, with no cohort trends:



As the hereditarian model explains group differences in terms of individual differences, it is reasonable to expect the black-white gap in small children to reflect shared environmental variation to a greater degree than in older children or adults. In particular, it may be that the “genetic nurture” (cf., Kong et al., 2018) provided by white parents is, on average, superior to that provided by black parents (and other caretakers), inflating the age-3 gap from what it would be if this shared environmental influence was absent. The results of the [Minnesota Transracial](#)

[Adoption Study](#), which studied children of various races adopted by affluent white parents, support this interpretation, with members of different racial groups scoring higher than expected at age 7 yet regressing to their racial means by age 17 (but note the small sample sizes):

Children's background	Number of Children	Age 7 Corrected IQ	Age 17 Corrected IQ
Non adopted, with two white biological parents	101	110.5	105.5
Adopted, with two white biological parents	16	111.5	101.5
Adopted, with one white and one black biological parent	55	105.4	93.2
Adopted, Asian or indigenous American parents	12	96.1	91.2
Adopted, with two black biological parents	21	91.4	83.7

The causes of the black-white gap at different ages could be investigated in twin studies or admixture studies. Genetic effects on the gap would be expected to increase and shared environmental influences decrease with age in twin models. In admixture analyses, the correlation between white ancestry and IQ in African-American children would be expected to become greater with age.

3.1.5. Biological gradient

Greater exposure should generally lead to greater incidence of the effect.

In the context of the hereditarian model of the black-white gap, the biological gradient refers to the expectation that in a racially admixed population, such as African-Americans, a greater amount of white ancestry should correspond to higher IQ. The fact that this constitutes a test of the hereditarian model was recognized long before there were data to conduct a proper test. For example, Jensen discussed the method in his response to criticisms of his *Harvard Educational Review* article in 1969:

Brazziel is quite correct in noting, for example, that the Negro population of the United States, like the white, is very far from being genetically or racially homogeneous. In fact, it is doubtful that any babies of pure African descent are being born in the United States today, unless they are born to African exchange students. But Africans, too, are genetically heterogeneous. A number of studies based on the differential frequencies of various blood groups in African and Caucasian populations have shown that, on the average, persons socially classified as American Negroes now have an admixture of 20 to 30 per cent Caucasian genes [...]. The percentage of Caucasian admixture varies greatly in various regions of the country, going from an average of below 10% in some Southern

states to above 25% in some Northern states. These figures can be estimated with considerable precision in large population samples, depending on the number of different blood groups and other genetic polymorphisms one is able to take into account. With these methods individuals, too, can be categorized by proportion of Negro-Caucasian admixture on a probabilistic basis. Possibly these same genetical techniques could provide a basis for more refined and accurate tests of hypotheses concerning racial differences in ability patterns. Since skin color is but poorly correlated with the percentage of Caucasian admixture, and because it may have social-environmental consequences, it could be statistically controlled in studies of the correlation between Negro-Caucasian admixture and measures of psychological characteristics.

–Jensen, 1969b

It is only very recently that data suitable for testing the relation of European admixture to IQ in African-Americans have become available. Kirkegaard et al. (2019) and Lasker et al. (2019) tested it in samples of black children and adolescents. The hereditarian prediction was borne out in both studies. Lasker et al. (2019) is particularly important, because it had a large sample of self-identified African-Americans (N=2,179). The paper reports various analyses, but the most straightforward to interpret, in my view, are Models 1 and 2 in Table 5:

Table 5. Regression analysis for European ancestry as a predictor of g among monoracial African-Americans with controls for skin color (Model 2), and SES (Model 3) added.

	Model 1			Model 1b			Model 2			Model 3		
Predictor	B	SE	β	B	SE	β	B	SE	β	B	SE	β
Intercept	-1.16	0.04	-1.01	-0.82	0.06	-0.997	-1.05	0.09	-0.996	-0.83	0.09	-0.995
EUR	0.78 ***	0.19	0.09 ***				0.83 ***	0.24	0.10 ***	0.67 **	0.23	0.08 **
Skin Color				-0.13 ***	0.04	-0.09 ***	-0.07	0.04	-0.05	-0.06	0.04	-0.04
SES										0.36 ***	0.03	0.28 ***
Adjusted R ²	0.007			0.006			0.013			0.082		
N	2179			1526			1526			1500		

Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. Model 1b shows the results with color as an alternative predictor. EUR = European ancestry. SES = socioeconomic status.

The models estimate the relation of white admixture to IQ in self-identified blacks with and without a control for skin color. The results indicate that about 60 percent of the IQ difference between blacks and whites can be attributed to genetic differences, with skin color exerting no significant influence on IQ independently of ancestry.

A caveat here is that the estimate has a rather wide confidence interval, and the analysis is also compatible with the inference that considerably less than 50 percent (or, at the other end, a full 100 percent) of the gap is genetic. However, the point estimate is similar to that in the earlier study by Kirkegaard et al. (2019), and when Lasker et al. widened the analytical sample to include biracial and white individuals, the estimates were similar or higher. Therefore, the best available estimates from admixture analyses are very consistent with predictions from the hereditarian model. Importantly, Lasker et al. also performed a measurement invariance analysis which showed that none of the one standard deviation black-white IQ gap in the sample could be attributed to test bias.[\[Note\]](#)

According to most philosophers of science today, strong falsification tests in the Popperian sense are not what science is about, and it is certainly true that a theory can always be made to accommodate unexpected findings. However, I would say that if there was no positive relation between white ancestry and IQ in African-Americans, the hereditarian model would be difficult to save. The fact that admixture analyses do now confirm this “risky” hereditarian prediction originally made half a century ago speaks strongly in favor of the model.

3.1.6. Plausibility

A plausible mechanism between cause and effect is helpful (but Hill noted that knowledge of the mechanism is limited by current knowledge).

Unlike environmental explanations of black-white IQ differences, the hereditarian model is rooted in a micro-level theory of causation, namely Mendelian genetics. Currently, genome-wide association studies are in the process of pinpointing genetic polymorphisms underlying IQ variation. It will take a long time to account for most or all of heritability using specific genetic variants, but much progress can be made by simply increasing sample sizes.

In his article on the three “laws” of behavioral genetics^[Note], Turkheimer (2003) argued that the “apparent victory of nature over nurture suggested by the first two laws is thus seen to be more methodological than substantive.” If there were “environmental twins”, with known proportions of environmental sharing, it would be straightforward to estimate how much shared experiences contribute to trait variation, analogously to how the known proportions of genes shared by twins enable the precise estimation of the magnitude of genetic effects. Thus, Turkheimer claims, the much greater tractability of genetic effects is solely due to the serendipitous existence of twins and other natural experiments that Mendelian inheritance continually creates.

I would say that the victory of nature is also substantive. The reason why the effect of genes on behavior is systematic is because genotypes are laid down at conception and remain pretty much invariant throughout life. You carry the same genes with you always and everywhere. In contrast, it is difficult to find systematic environmental sources of variation in a given broad population because, except for major physical trauma, environmental effects are fleeting and will exert a diminishing influence over time unless continuously reinforced. Environmental effects are of an infinite variety and provenance, unlike genes, all of which can be physically located and enumerated. The reason why environmental effects overwhelmingly appear to be akin to chance, noise, or randomness is not just methodological. One cannot recruit samples of “environmental twins” because, within a broadly defined population, very few, if any, people share non-genetic effects on IQ to anywhere near the extent that DZ twins, let alone MZ twins, share their genetic effects.

Environmental explanations of black-white differences assume the existence of constant and systematic non-genetic influences that push apart the trait value distributions of the two races often living side by side. Given what we know about the causation of individual differences, the discovery of many such systematic influences is unlikely. The constancy of black-white differences means that they have the qualitative nature of a genetic rather than an environmental phenomenon.

3.1.7. Coherence

The cause-and-effect interpretation of our data should not seriously conflict with the generally known facts of the natural history and biology of the disease.

A virtue of the hereditarian model, as formulated by Jensen, is that it seeks to explain group differences using the same variables that are used to explain differences between individuals. The hereditarian has no need to postulate the existence of a separate group-differences realm where the rules that you would use to analyze differences between individuals do not apply. In contrast, environmentalists believe that within-group heritability is uninformative about between-group differences. A metaphor about growing seeds in different environments is often used in this context. Herrnstein & Murray (1994, p. 298) described this environmentalist argument in this way:

That a trait is genetically transmitted in individuals does not mean that group differences in that trait are also genetic in origin. Anyone who doubts this assertion may take two handfuls of genetically identical seed corn and plant one handful in Iowa, the other in the Mojave Desert, and let nature (i.e., the environment) take its course. The seeds will grow in Iowa, not in the Mojave, and the result will have nothing to do with genetic differences.

Stephen Jay Gould (1996, p. 186–187) wrote along similar lines:

[V]ariation among individuals within a group and differences in mean values between groups are entirely separate phenomena. One item provides no license for speculation about the other.

[...]

Within- and between-group heredity are not tied by rising degrees of probability as heritability increases within groups and differences enlarge between them. The two phenomena are simply separate.

Similarly, Eric Turkheimer has [argued](#) that while heritability estimation within groups is legitimate, there is no scientific basis for estimating the heritability of group differences:[\[Note\]](#)

There is no such thing as a “group heritability coefficient,” no way to put any meat on the speculative bones about partial genetic determination.

To avoid any misunderstanding, it should be stressed that behavioral genetics does not study within-individual determinants of phenotypes. It studies determinants of differences between individuals, and behavioral genetic parameters are meaningful only with respect to some population. Behavioral genetics is *not* an [idiographic](#) science, and it does not seek to discover the molecular mechanisms involved in the development of a particular phenotype. The focus of behavioral genetic analysis is in the unbiased estimation of causal genetic and environmental effects on differences between individuals and groups.

What separates hereditarians from environmentalists like Turkheimer is not that the former focus on populations and the latter on individuals. Both study populations. The difference is that whereas hereditarians treat different races (within a single nation, at least) as members of the same population in the sense that they are subject to the same causal forces, environmentalists erect insurmountable barriers between races, in effect arguing that “black is black, white is white, and never the twain shall meet.” Ironically, it is in environmentalist explanations that races are reified into separate essences so that even when members of different races share the same social world, they always experience it in very different ways. Given the persistence of the IQ gap across generations, socioeconomic classes, and geographic locations, environmentalists are, in order to explain the gap, forced to postulate an extraordinarily tenacious and rigid linkage between racial identity and one’s experience of the world.

3.1.8. Experiment

“Occasionally it is possible to appeal to experimental evidence”.

In non-human research, the genetic basis of any individual or group difference can be established in controlled breeding experiments. Experimental manipulation of most genetic and environmental factors pertinent to IQ is not feasible in humans, and our behavioral complexity means that results from animal experiments are of little value for the study of human behavior, least of all IQ variation. However, once genetic engineering of humans, perhaps in the form of embryo selection based on polygenic scores, gets truly going, experimental methods may, in some limited sense, come within the reach of human behavioral genetics, too. More important, however, is the way that genetic engineering will force the issue of the population’s genetic quality to public attention and debate.

3.1.9. Analogy

The use of analogies or similarities between the observed association and any other associations.

In its tenacity across a wide variety of contexts, the black-white IQ gap resembles a phenotypic

difference with a genetic rather than environmental etiology. Environmental influences on IQ tend to be much more fickle and fleeting than genetic ones. Black-white IQ differences have a systemicity to them that argues against environmental causation.

Psychometric evidence points to genetic causes, too. As discussed in a previous section, black-white IQ differences are psychometrically quite dissimilar to IQ differences between people from different generations (or countries, cf., Wu et al., 2007). Environmental differences must be invoked to explain the Flynn effect, but those same environmental differences cannot be behind the black-white gap. This is part of a general pattern where black-white test score differences have been found to relate to psychometric parameters in a manner similar to that of strongly genetically influenced biological variables, and quite differently from how environmental variables are related to the same psychometric parameters. For an elaboration of this argument, see [here](#).

3.2. Conclusion

Hill (1965) stressed that his criteria do not establish any kind of deductive basis for causal inference, but rather that they help decide on the best-supported answer to the question at hand:

Here then are nine different viewpoints from all of which we should study association before we cry causation. What I do not believe – and this has been suggested – that we can usefully lay down some hard-and-fast rules of evidence that must be obeyed before we can accept cause and effect. None of my nine viewpoints can bring indisputable evidence for or against the cause-and-effect hypothesis and none can be required as a *sine qua non*. What they can do, with greater or less strength, is to help us to make up our minds on the fundamental question – is there any other way of explaining the set of facts before us, is there any other answer equally, or more, likely than cause and effect?

The correct explanation of black-white IQ differences, like all scientific explanations, must rely on abductive inference—arriving at the most likely answer given all the evidence. It is not a mathematical or logical problem that can be solved so that not a shadow of a doubt remains. To argue that because the problem is complex, or because we cannot be 100 percent certain of the answer, we should refrain from trying to answer it at all is not an argument against race and IQ research in particular. It is an argument against *all* scientific inquiry. Nor is the race and IQ question an exceptionally hard problem in the greater scheme of things. The mechanisms described by Mendel's laws are nature's way of generating vast amounts of high-quality quasi-experimental data on individuals in their natural settings, enabling causal hypotheses to be tested much more easily and credibly than is normally possible in the social sciences.

Appeals to “irreducible complexity” and the like would be more tolerable if they came from a place of genuine skepticism and elevated epistemological standards. In practice, however, the same people who assert that genetic causation of behavioral differences in humans is too difficult a problem to solve will readily attribute the same differences to environmental causes, as if similar complexities (and others as well) did not obtain for non-genetic causation. [Beware of isolated demands for rigor](#). For example, in [this recent article](#), four geneticists first argue at length that establishing whether genetics is involved in group differences is far too hard a problem to answer at the moment (they ignore all the actual hereditarian arguments, such as those presented in this post). Having done that, they cast their demands for methodological rigor aside and confidently proclaim that “any apparent population differences in IQ scores are more easily explained by cultural and environmental factors than they are by genetics.”

To move the debate forward, I would love to see non-hereditarians formulate their models in more detail, with more attention paid to testable implications, psychometric and behavioral genetic details, and heuristics like Hill’s criteria. This would enable more involved comparisons between genetic and environmental explanations. A back-and-forth dialectic between different scientific “camps” appears to me to be essential to scientific progress, as long as it is done in good faith.

4. The past, present, and future of race and IQ

4.1. Private truths

In 1987, *The American Psychologist* published a comprehensive, anonymously conducted survey of the views of research psychologists and other behavioral scientists (N=661) on IQ and related topics (Snyderman & Rothman, 1987; this survey was later expanded into a book: Snyderman & Rothman, 1990). Surprisingly, the survey revealed that hereditarian views on IQ were widespread among social scientists. Not only was there a widely shared conviction that IQ is substantially heritable and that social class differences in IQ are at least partly genetic, but also no less than 46% of the surveyed researchers stated that genetic factors explained at least some of the black-white IQ gap. In contrast, only 15% of the respondents subscribed to a purely environmental model of the black-white gap.

The survey therefore revealed that among the respondents alone, there were 304 researchers with at least somewhat hereditarian views on race and IQ. On the other hand, as anyone familiar with the relevant scholarly debates in the decades since Jensen’s 1969 paper can attest, there were nowhere near 304 academics who had publicly admitted to such hereditarian beliefs. The ranks of open hereditarians were thin, their numbers countable with the fingers of two hands. Many, many more professed belief in something like the blank slate in academic papers and public declarations. For example, a “Resolution against Racism” signed by over 1,000 American academics was published in the *New York Times* in 1973. According to Segerstrale (2000, p. 33), the statement “declared that all humans have been endowed with the same intelligence”,

condemning the research of Jensen and others as “both unscientific and socially pernicious.”[\[Note\]](#)

The fact that hereditarianism with respect to racial differences was publicly marginal and privately mainstream among research psychologists is a good example of what Timur Kuran has called *preference falsification*: the tendency of people to mask their true beliefs in order to conform to what is socially acceptable. It is also an example of how a small but intransigent minority of researchers can make alliance with powerful political forces and hijack a scientific discourse. Jensen won a Pyrrhic victory of sorts in the debates that raged after his 1969 article: he appears to have convinced many intellectually, while his opponents convinced few, but in terms of public influence and recognition the latter carried the day. (See [here](#) for some evidence that the arguments of Jensen and the few other out-of-the-closet hereditarians did in fact shift the views of academics towards a more hereditarian outlook after 1969.)

In the years since Snyderman and Rothman’s survey, the most significant collective endorsement of hereditarianism has been “Mainstream Science on Intelligence”, a remarkable document penned by Linda Gottfredson in 1994. Birthed by the *Bell Curve* controversy, the statement appeared in the *Wall Street Journal* and was signed by 52 researchers in intelligence, behavioral genetics, and related fields. Its defense of the role of heredity in individual differences in IQ was full-throated, while its discussion of the causes of the black-white IQ gap was more oblique and hedged (“[m]ost experts believe that environment is important in pushing the bell curves apart, but that genetics could be involved too”). This ‘hereditarianism lite’ both affirmed the scientific validity of the racist approach, and conferred plausible deniability to individual signatories. Even so, the statement must have appeared alarming to people adamant on the thesis that any cognitive deficiency of blacks must, at a deeper level, be a reflection of a deficiency not of blacks but of whites. The prominent leftist behavioral geneticist Eric Turkheimer voiced some of these concerns in a 1997 piece where he called for an intellectually serious “psychometric left” to counter the influence of the predominant “psychometric right” (Turkheimer, 1997).

However, in retrospect it appears to me that the *Mainstream* statement may have been more of a last stand for racially informed hereditarian thinking. The generation of scientists who signed the statement in 1994 mostly came of age academically before blank slatism with respect to race differences stiffened into orthodoxy in the social sciences. A useful illustration of the generational shift is that segregationist views did not prevent psychologist Henry Garrett (1894–1973) from being both President of the American Psychological Association (1946) and Chair of Psychology at Columbia University (1941–1955). In contrast, Arthur Jensen (1923–2012), a much greater scientist than Garrett, had an uncertain position within the wider psychological research community: he was highly influential but also something of a pariah who was certainly never in a position to exert institutional influence over psychology.

After Jensen’s time, a handful of researchers have continued his project, but none have his scholarly prestige, and it seems indisputable that the marginalization of racist research has intensified. For example, while Jensen’s research was frequently published and debated in high-

impact general psychology journals, these days the subject is usually relegated to small specialist journals. Getting involved in racial differences research is a precarious pursuit for a young academic these days, as recently shown by the [case of Noah Carl](#). His engagement with race research was tangential yet it weighed heavily in the decision to fire him from his position at the University of Cambridge. Not coincidentally, some of the most interesting work on race and IQ is now being done by independent researchers with no institutional backing.

Researchers may be cajoled into silence when the dominant ideology of the day tolerates dissent poorly, but that does not mean that they have actually been persuaded of the veracity of the politically correct viewpoint. However, what if students never even encounter views that challenge the conventional wisdom? What if the published literature does not reflect the actual convictions of scientists because of self-censorship? An interesting question is if a survey like Snyderman and Rothman's would today show that blank slatism with respect to race has won the day at long last. If the hereditarians of one generation voice their true views only *sub rosa* and not in their publications or teaching, the next generation of researchers may well take the public, "exoteric" views of their teachers at face value.

Some evidence against the conjecture that hereditarianism is on the wane comes from a study of Rinderman et al. (2020). Anonymously surveying authors who had published in journals associated with intelligence research, they found that closet race realism is still prevalent in psychology: the respondents' mean estimate for the heritability of the black-white gap was 49 percent, and only 16 percent of the respondents put the estimate at zero. However, a different sampling frame and a low response rate in the new survey preclude a direct comparison with the old one.

4.2. Who has the relevant expertise?

In discussions of race and IQ, what is often forgotten is that it is primarily a *psychological* and *psychometric* question, and only secondarily a genetic or biological one. A common misperception is that geneticists or even neuroscientists are best positioned to resolve this old dispute. After all, the contention is ultimately about how genetics contributes to population differences in the structure and function of the brain. Nevertheless, this line of thinking is roughly akin to telling someone with a toothache to consult a mineralogist rather than a dentist. Geneticists and neuroscientists do not study intelligence, lacking even the rudiments of the expertise needed to assess the results of cognitive tests. It is not reasonable to expect them to supply answers to questions that they rarely, if ever, ask.

If you do not know how to interpret the results of mental tests, you have no way of evaluating the evidence at the center of the dispute. Psychometric modeling provides information that is independent of and complementary to that provided by genetic (and, sometimes, neuroscientific) methods. This crucial source of causally relevant information is not available to those who approach the problem from a non-psychological perspective. In fact, if you do not have

psychometric evidence for the unbiasedness of the tests used, all of your conclusions about between-group test results are on a very shaky ground.

As should be apparent from the default model described earlier, hereditarian research on race and IQ is primarily concerned with *proximate* genetic explanations. The question at issue is what *is*: Is there a significant genetic difference in IQ between races X and Y, and if so, what is its magnitude? In contrast, *ultimate* genetic questions (why did these genetic differences arise? what kind of evolutionary forces were involved?) are, in my opinion, at best of secondary importance. Ultimate genetic explanations are always less certain than proximate ones. We may be able to establish with high confidence the genetic basis of a phenotypic difference between contemporary populations even as the ultimate explanation of the difference—such as specific, long-standing selective pressures—remains elusive. This is because the same outcome could have originated from many different evolutionary histories.[\[Note\]](#)

The theory and methods of proximate genetic explanation of human behavioral variation—that is, behavioral genetics—were principally developed by psychologists who often built on the insights of quantitative geneticists, in particular animal breeders. The expertise of an animal breeder overlaps somewhat with that of a genetically oriented psychologist, but a deep understanding of controlled breeding is of limited utility in the (by necessity correlational) study of quantitative traits in humans. Moreover, measuring IQ in humans poses rather different challenges and requires different kinds of expertise than measuring, say, milk yield in cattle. Other kinds of geneticists, whether “wet-lab” types or theoretical population geneticists, or anything in-between, have even less competence to respond to questions of primary importance in race and IQ research. The typical molecular geneticist, of course, does not know more about behavioral genetics or IQ than your average man on the street. Population geneticists, in their turn, specialize in the study of changes in allele frequencies over time, but that expertise alone gives no special insight into the race and IQ problem. It is not possible infer the existence or absence of a genetic difference from purely theoretical considerations. There is no substitute to getting into the weeds of the problem at hand, which involves grappling with both its genetic and measurement particulars.

As for neuroscience, it is a rather primitive field at the moment. At best, neuroscience is capable of supplying correlational information about what brain regions appear to be involved, in some unspecified manner, with a behavioral variable, such as IQ. Even this modest ambition is often thwarted by the cottage industry nature of the field where numerous small labs have produced research literatures largely consisting of false positives (Button et al., 2013). Furthermore, neuroscientists do not receive any measurement training, and seem to carry out their business under some sort of naive operationalism. In any case, the definitions of concepts like intelligence come from their behavioral associations, not any neurobiological correlates that they may have. Neuroscience can, at best, hope to discover brain structures and functions that are isomorphic to IQ. The (rhetorical) question then is why one would try to study race and IQ using the imperfect correlates of IQ in the brain that contemporary neuroscience may discover, rather than studying

IQ itself, with its unparalleled attendant literature on validity, measurement invariance, heritability, and so on.[\[Note\]](#)

In short, to understand group differences in IQ, one must have a strong familiarity with certain psychological research traditions. I am not saying that one should be a psychometrician or behavioral geneticist by training to study this problem—disciplinary gatekeeping is the least of my concerns—but usually it takes at least a few years of grappling with the literature and the data to really grasp the (literal and proverbial) parameters of this debate. There is simply no research tradition outside of psychometrics and behavioral genetics that provides tools for answering these questions.

4.3. Why study it?

Black-white disparities—educational, occupational, financial, familial, those related to criminal justice and health, and many others—are the obsessive focus of enormous amounts of academic research and media coverage in America. In this discourse, hereditarian explanations are almost always either entirely ignored or glibly dismissed.

A typical example of how the cognitive gap between whites and blacks is covered in the prestige media is [this story](#) by Laura Meckler than ran in the *Washington Post* last October. It is about Shaker Heights, an affluent suburb of Cleveland, OH, which prides itself on its generations-long commitment to racial integration. Yet, somehow, that has not dented the black-white achievement gap:

But the story of Shaker Heights shows how moving kids of different races into the same building isn't the same as producing equal outcomes. A persistent and yawning achievement gap has led the district to grapple with hard questions of implicit bias, family responsibility and the wisdom of tracking students by ability level.

While the *WaPo* article does not mention it, the black-white test score gap in Shaker Heights public schools is around 1.6 standard deviations, which is the third largest gap in the nation according to the comprehensive study by Reardon et al. (2019). Only two other cities—Chapel Hill, NC and Berkeley, CA—both also wealthy and liberal, had larger gaps between white and black students. While Shaker Heights may be unusual in terms of both the size of the gap and the lengths to which the residents have gone trying to eliminate racial inequities, real or perceived, in a qualitative sense the story is entirely generic. It could be about any place in America. As Reardon et al. showed, there is a black-white achievement gap, favoring whites, in all 2,300 school districts in the country with at least 15 students of both races. They also noted that black students tend to be enrolled in districts with higher per-pupil spending than white students, which

probably reflects pervasive attempts to close achievement gaps. The *WaPo* article tries to find the causes of a universal phenomenon in the specific circumstances of a particular place. [\[Note\]](#)

The main framing device in Meckler's article is the experiences of Olivia, a black teenager from Shaker Heights who felt disrespected by a white teacher. A community meeting was eventually called to address her concerns. So what did the teacher do? Meckler writes:

The racial tension coursing through the packed auditorium last November traced back to a tense exchange between Olivia and a veteran AP English teacher, Jody Podl, six weeks earlier. Olivia had been dozing in class, playing with her phone. Now, her first big assignment of the year was late. The teacher had admonished and embarrassed Olivia. Olivia's mom fired off a three-page complaint, suggesting racism and charging bullying. The district put the teacher on leave to investigate.

We have come from Bull Connor with his fire hoses and attack dogs to a white teacher being hung out to dry for doing her job while failing to properly heed the *amour-propre* of a slacking black student. The blank slatist is forced to regard such momentous transformations in racial attitudes and power relations as irrelevant details with no wider significance. A Jody Podl is just Bull Connor with a polite face, enforcing white supremacy through such nefarious acts as demanding that her students complete their homework. How else could one explain the stubborn refusal of reality to conform to one's expectations even after decades of good-faith efforts? Meckler describes sympathetically the concerns of the aggrieved blacks and perplexed whites of Shaker Heights. She does not have much to say about causes, except that "economics" is involved and that "gaps are also social", as reflected in the disproportionately white composition of parent-teacher associations.

Meckler's article represents a familiar genre of writing. Numerous articles finding the same achievement gap in any number of different places have been published. These articles usually have a surprised, breathless tone to them, as if they were reporting something unprecedented. However, journalists on the education beat cannot be blamed too hard for churning out these silly stories. The scholarly community approaches the topic with the same kind of politically convenient incuriosity. For example, in the aforementioned study that found black-white gaps absolutely everywhere, Sean Reardon and colleagues simply assert, without argument, that the hereditarian thesis is "supported by no credible theory or evidence", and go on to list various feeble environmental hypotheses, drawing on the correlational, thoroughly genetically confounded data that they had access to. When academic experts on education have only such pabulum to offer, it is difficult for journalists to broach the topic from a more realistic and balanced perspective. [\[Note\]](#)

Despite all the resources and accolades available to purveyors of environmental and cultural determinism in the IQ controversy, its results remain unimpressive. It offers piecemeal, *ad hoc* explanations that are rarely in any way integrated with important findings from relevant disciplines, especially psychometrics and behavioral genetics. There is a distinctively non-progressive character to this endeavor, with each successive generation of researchers discovering the gaps anew and proposing variants of the same explanations, drawing on correlational, genetically confounded analyses, or small, unlikely-to-replicate experiments. The causal hypotheses proffered appear to be often inspired not by plausibility but by a need for a certain kind of moral narrative. As a consequence, the legacy of slavery and segregation has assumed an ever more powerful explanatory role in discussions of black-white differences even as the era of the legal subjugation of blacks has receded into a distant memory and demographic changes have made the black-white dichotomy in many ways socially anachronistic.

It is a matter of basic intellectual and moral hygiene that blank slate proponents not be allowed to monopolize a discourse that takes such a large space in the contemporary imagination. For a hereditarian to keep silent is akin to unilateral disarmament during a hot war. Heredity is the primary systematic source of individual differences, so the suggestion that it should be *a priori* ruled out as an explanation of racial differences is preposterous. The hereditarian program provides ready explanations to many of the gaps that so puzzle the environmentalist, together with a strong framework to test them. Research on IQ is especially relevant because it is perhaps the primary axis of behavioral differentiation between blacks and whites in America; if the IQ distributions of whites and blacks were not so different, the “race question” would, I believe, lose most of its steam. (Racial differences in anti-social behavior and crime are another important axis of differentiation, but it is partly a function of IQ, too.)

Contra the frequently voiced concern about the dangerous societal implications of research on race and IQ, I am very skeptical that it could have any great political or social consequences. I believe that when it comes to questions of great importance to society, whatever the best science says will always be very easy to ignore or twist beyond recognition if needed; the alignment of science and politics, if it happens, is more accidental than causal. Nor do day-to-day interactions between people change based on what the latest peer-reviewed research says. If racist thinking and praxis reached their peak strength around, say, the early 20th century in conjunction with the heyday of European colonial empires and white supremacist policies in various European offshoot countries, it was not because the theoretical and empirical case for racism had in those days achieved great new heights. Nor was hereditarian and racist thinking made anathema in the 1960s and 70s as a result of any scientific discoveries. Science is a thin reed to cling to when powerful social and political currents are at work.

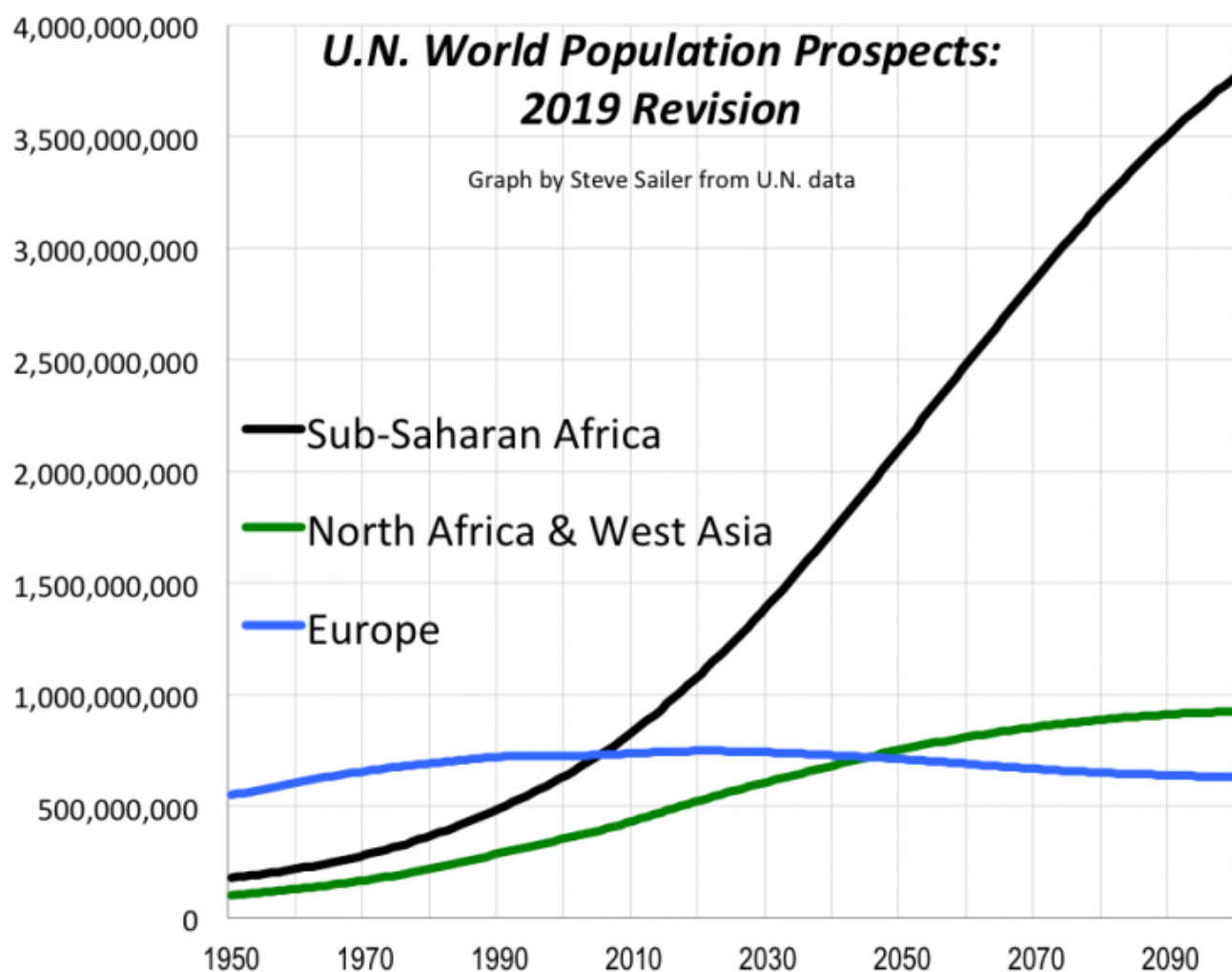
From this perspective, the frequent moral panics about racist thinking and attempts to drive it out of academia betray a very exaggerated belief in the power of scientific arguments to effect social change. It reflects a narcissistic conviction by academics that their work has great societal significance. Equally far-fetched is the idea that a resurgence of race realist research would

somehow rectify demographic and other trends that so alarm conservatives in contemporary Western societies.

The reason for researching race and IQ is therefore to understand, predict, and anticipate how people and societies function and develop. The research may persuade you, personally, to adopt certain political views and advocate for them, but never count on others who do not share your political instincts to be similarly persuaded. It is quite possible to win the intellectual argument by convincing those who are convincible by intellectual arguments, yet win no influence at all in society at large—this is what seems to have happened to Jensen, as per Snyderman and Rothman's survey. If you want to change society, scientific research may not be the best investment of your time.

The current century will offer the racially informed observer plenty to look at. In many ways, the future is already here. We can see the familiar, persistent racial disparities in the educational achievement of today's schoolchildren in America. The demographics of human capital for the rest of the century will offer no surprises.

Globally, the human biodiversity issue of the 21st century is the demographic decline of populations with high human capital (especially in Europe, East Asia, and North America) and the rapid demographic expansion of populations with low human capital (especially in sub-Saharan Africa). "The world's most important graph" by Steve Sailer is a good visualization of this problem:



Race realist research has thus far mostly focused on black and white Americans, reflecting both the early salience of the “race problem” in America and the availability of high-quality data. More and better research from an international perspective is needed. Mostly descriptive research such as Richard Lynn’s work on international IQ differences should be extended with causally more sensitive study designs. The recent paper by Kirkegaard et al. (2018) shows one way forward.[\[Note\]](#) Racist research will likely continue to face vigorous opposition, given that, as a result of the immigration policies of recent decades, most Western societies are now facing problems in the management of increasingly multiracial populaces and labor forces.

I close with some of Arthur Jensen’s reflections on his famous 1969 article and the research program that it started, published in 1978 in the context of the article being recognized as a “Citation Classic”:

My unrelenting research in the so-called 'IQ controversy,' for example, has resulted, over the years, in the loss of the friendship of a number of my colleagues; in near-riotous demonstrations by student activists at many colleges where I have been invited to speak; and in last-minute cancellation of invited lectures, vilification in student newspapers and leaflets, and physical threats to me and my family, occasioning the need for police protection, even as recently as a month ago, and as far away as Australia. One may imagine subtler penalties, too, such as the loss of academic status and respectability, but this is more difficult to assess. It does not worry me perhaps as much as it should. The fact that I am not only alive and well, but reasonably happy and unstintingly carrying on my research on all aspects of human intelligence will no doubt be attributed to personal eccentricity. But I hope it will also be encouraging to others. From my experience I can say that, in the long haul, the consequences of sticking your neck out when you think you should, are not too bad. It is an exercise in conscience and self-respect, in which neither suffers, given the faith that the scientific pursuit of the currently most tabooed question will prove worthwhile to humanity.

—Jensen (1978)

Notes

1. The terms *racialism* and *race realism* are used in this post to refer to the idea that genetic variation is an important cause of differences between human populations not only in physical but also in psychological and behavioral characteristics.
2. This philosophy of measurement is known as *construct validation* in psychometrics. Cronbach & Meehl (1955) is the classic article on the idea.
3. 40% would be a typical proportion of variance explained by *g* in contemporary test batteries.
4. The *g* factor appeared to be much more of a formative (statistical) variable in Jensen's thinking in 1969 than it was in his later work where the invariant and reflective (causal) nature of *g* was emphasized. My defense of *g*, which draws heavily on Jensen's mature thinking, can be read [here](#).
5. Schmidt & Hunter (2004) and Salgado et al. (2003) provide modern meta-analytic estimates of the associations between IQ and various aspects of job performance.
6. Warne et al. (2013) analyzed several datasets, concluding that the IQ distribution does not

seem to have a “fat” right tail.

7. I wrote about newer research on the stability of IQ [here](#).

8. It seems improbable that a group of equally eminent geneticists would today issue a statement supporting the feasibility of such eugenic measures, even though the inference is, of course, even more firmly supported today than it was in 1967.

9. Jensen uses a value of .60 for the spousal IQ correlation. It is clearly too high a value in light of more recent research. Bouchard and McGue (1981) found an average correlation of .33 in sixteen samples, while the correlation was 0.35 in the study by Keller et al. (2013). Jensen’s empirical estimates are therefore probably inflated, but the principle that (positive) assortative mating increases variance is correct.

10. Burt’s results regarding the heritability of IQ and certain other things came under strong suspicion of research fraud soon after his death in 1971, two years after Jensen’s article was published. For example, it is not clear whether he actually had data on nearly as many twins reared apart as he claimed. However, even if Burt’s results were fraudulent, in retrospect this had little effect on the big picture because his numbers are well in line with those of many other researchers whose honesty is not under question. It would be useful to analyze the credibility of Burt’s data with modern methods such as the [GRIM](#) and [SPRITE](#) tests.

11. In modern reviews of research on twins reared apart, Burt’s results are disregarded. There are five generally accepted studies of IQ in MZ twins raised apart. The weighted average correlation in them is .75, without adjustment for unreliability (Bouchard, 1997).

In my view, studies of MZ twins reared apart, while conceptually attractive, cannot provide a firm basis for inferences about the magnitude of heritability. This is because there are only a handful of small studies of such twin pairs, and they are based, by necessity, on unusual convenience samples rather than random samples. The real basis for the heritability estimation of human traits is the comparative analysis of MZ and DZ pairs reared together, or the classical twin design. The samples used in classical twin studies are large and often nationally representative. The only serious conceptual criticism of the classical design concerns the assumption that the environments of MZ and DZ pairs are equal, and thorough assessments of the assumption indicate that at worst it adds random error, rather than systematic bias, to estimates (Conley et al., 2013; Felson, 2014; Barnes et al., 2014). See also [my post](#) on how sex differences (or lack thereof) can be used to test the equal environments assumption.

While the genetically informative data available to Jensen in 1969 was much less than optimal—certainly so in the case of Burt’s data—his conclusion that genetic effects on IQ are large and shared environmental effects modest is strongly borne out by later research. A broad sampling of results from recent studies on the heritability of IQ is shown in the table below. It is

not a comprehensive meta-analysis, but I believe that it gives a reasonable overview of typical estimates.

The studies are in the order of publication, but do note that sample recruitment dates vary widely regardless of publication date, and that samples overlap between some studies. The estimates reported are from the authors' preferred models. Some estimates were not directly available in all studies, and were calculated from twin correlations using the Falconer method. "Classical twin" refers to comparisons of MZ and DZ twin reared together, whereas "extended twin-family" refers to various study designs that use other types of siblings and relatives besides reared-together MZ and DZ twins. The residual variances not accounted for by either heredity or the shared environment are not shown in the table but should be straightforward to compute.

Study	Country	N	Study design	Variable	Age	h^2	c^2
Silventoinen et al. (2006)	Netherlands	400	Classical twin	IQ composite	5	24%	52%
Silventoinen et al. (2006)	Netherlands	400	Classical twin	IQ composite	7	39%	30%
Silventoinen et al. (2006)	Netherlands	300	Classical twin	IQ composite	10	79%	0%
Silventoinen et al. (2006)	Netherlands	300	Classical twin	IQ composite	12	83%	0%
Silventoinen et al. (2006)	Netherlands	400	Classical twin	Raven's Matrices	16	61%	0%
Silventoinen et al. (2006)	Netherlands	400	Classical twin	IQ composite	18	83%	0%
Silventoinen et al. (2006)	Netherlands	400	Extended twin-family	IQ composite	26	84%	0%
Silventoinen et al. (2006)	Netherlands	600	Extended twin-family	IQ composite	50	83%	0%
Read et al. (2006)	Sweden	650	Classical twin	IQ composite	~79	68 %	0%
Friedman et al. (2008)	US	600	Classical twin	IQ composite	16	69%	16% (ns)

Study	Country	N	Study design	Variable	Age	h^2	c^2
Cesarini (2010)	Sweden	155000	Extended twin-family composite	IQ		18	71% 14%
Lee et al. (2012)	Australia	400	Classical twin composite	IQ		71	74% 0%
Vinkhuyzen et al. (2012)	Netherlands	900	Extended twin-family composite	IQ		47	82% 0 %
Beaver et al. (2013)	US	8300	Extended twin-family composite	IQ		4	45% 27%
Beaver et al. (2013)	US	8300	Extended twin-family composite	IQ		7	55% 19%
Beaver et al. (2013)	US	4200	Extended twin-family	Vocabulary		16	47% 23%
Beaver et al. (2013)	US	4200	Extended twin-family	Vocabulary		22	49% 20%
Keller et al. (2013)	US & Australia	7900	Extended twin-family composite	IQ	Varies		78% 8%
Krapohl et al. (2014)	UK	13000	Classical twin composite	IQ		16	58% 4%
Panizzon et al. (2014)	US	1200	Classical twin	Latent g factor		55	86% 0%
Bates et al. (2016)	Australia	2300	Classical twin composite	IQ		16	86% 8% (ns)
Engelhardt et al. (2016)	US	800	Classical twin	Latent g factor		11	77% 12% (ns)
Turkheimer et al. (2017)	Norway	2200	Classical twin composite	IQ		18	66% 18%
McGue et al. (2017)	US	1900	Extended twin-family composite	IQ	~15–17		61% 17%
Mollon et al. (2018)	US	4700 (whites)	Genomic composite	IQ		14	72% N/A
Mollon et al. (2018)	US	1900 (blacks)	Genomic composite	IQ		14	61% N/A
Gottschling et al. (2019)	Germany	1000	Classical twin composite	Non-verbal composite		11	53% 27%
Gottschling et al. (2019)	Germany	1100	Classical twin composite	Non-verbal composite		17	81% 4%
Gottschling et al. (2019)	Germany	1000	Classical twin composite	Non-verbal composite		23	67% 21%
Hur & Bates	Nigeria	3200	Classical twin	Raven's Matrices		15	31% 25%

Study	Country	N	Study design	Variable	Age	h^2	c^2
(2019)							
Hur & Bates (2019)	Nigeria	3200	Classical twin	Vocabulary	15	40%	28%
Median estimate						68%	12%

N refers to the approximate number of individuals in a study, while **Age** is the average age of study participants. h^2 is heritability and c^2 is the shared environment.

The median heritability in these studies is 68 percent, while the shared environment accounts for a median of 12 percent. There is a moderate amount of between-study heterogeneity in the estimates, which is probably mainly due to differences in participant age (heritability is lower and shared environmentality stronger in children), and differences in the reliability and factor composition of the tests used (scores from longer and more factorially diverse tests are more heritable). 80 percent is a reasonable estimate for the heritability of IQ in an adult sample from a wealthy country when IQ is measured with high reliability using a multiability test battery.

12. As noted by Jensen, research designs using measured environments and adoptee data depend on the assumptions that the measurement is fine-grained enough and that adoptees are randomly placed in families. These assumptions, which pull estimates of the effect of the environment in opposite directions, may be difficult to meet in practice.

13. Scholastic attainment is moderately to strongly influenced by the shared family environment. For example, the meta-analysis by Branigan et al. (2013) found that the average heritability of academic attainment in ten countries, measured as years of schooling completed, was 40% while the shared environment explained an average of 36% of the variation. A shared environmental component of that magnitude (even if the estimate is somewhat inflated by the failure to model assortative mating in the meta-analysis) is unusually large in behavioral genetics. Most estimates of the influence of the childhood family environment in adults, regardless of phenotype, are in the range of 0–20%, and they are frequently not statistically significantly different from 0.

Nevertheless, Jensen's claim that up to 60% of the population variation in scholastic performance can be explained by the family environment does not find support in modern datasets, even if one adjusts for measurement error as he does. Generally, educational achievement is less heritable and the shared environment more important when attainment is measured using imprecise variables such as "years of education completed" or tests evaluating narrow skills, while heritability is higher and the shared environment less important when achievement is measured using tests evaluating broadly defined skills. A meta-analysis by de Zeeuw et al. (2015) of test-

based educational achievement in 61 studies from various countries found the following average results, with the shared environment explaining clearly less variance than heredity:

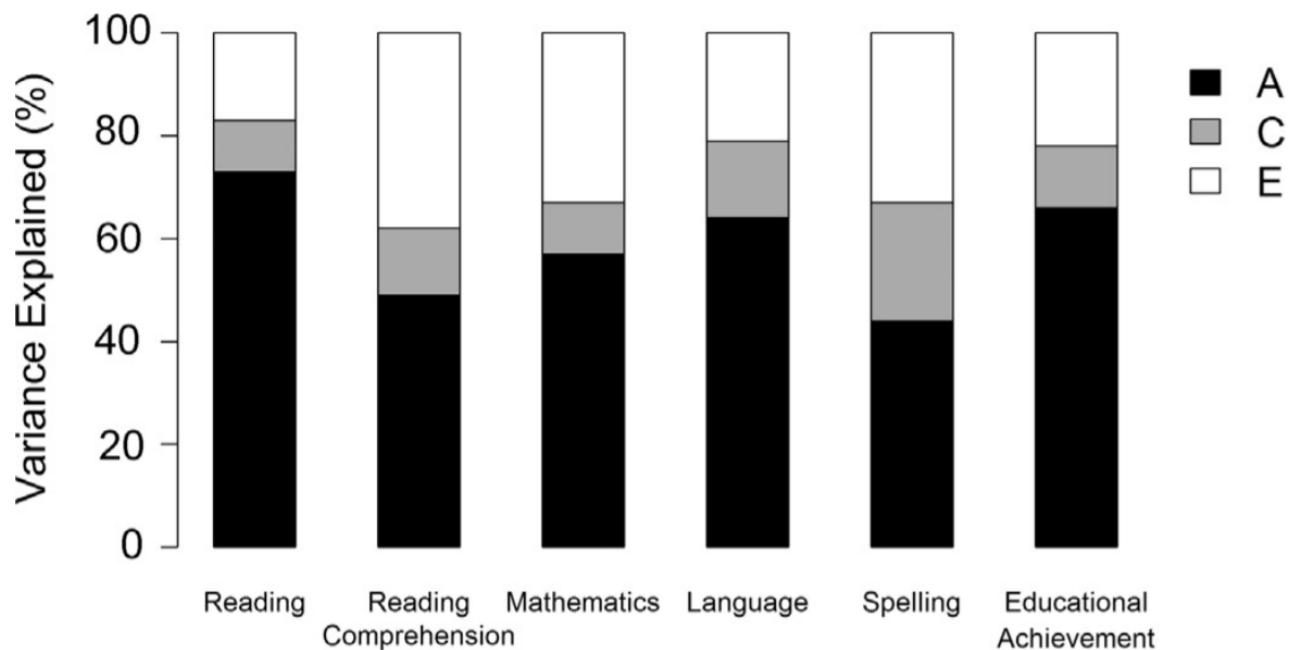


Fig. 2. Estimates for the genetic (A), common environmental (C) and unique environmental (E) effects from the meta-analysis of educational achievement.

It is not clear why Jensen found the shared environment to be much more important for scholastic achievement than modern research does. It may be that, due to homogenization of educational opportunities in the last 50 years, the importance of the family environment has decreased.

Alternatively or additionally, the relatively scarce data available to Jensen may have led him astray.

14. Jensen's belief that non-cognitive skills are a promising target for intervention is shared by many of today's influential education policy researchers, such as the economist James Heckman. The disappointing track record of attempts to boost IQ shifted the focus of reformers from intellectual ability to non-cognitive skills, but it remains unclear whether this line of thinking can escape the bite of [Rossi's metallic laws](#) any better than cognitive interventions. The belief in the ability of preschool programs and the like to permanently change personality and motivation rests on evidence from old, small, and never replicated experiments like the [Perry Preschool Program](#).

15. "Culturally disadvantaged" was a term of art in older social science literature. It seems to have referred to black and other low-achieving minority children and to poor whites.

16. There is a cloud of suspicion hanging over Rick Heber's research. This is because he later led an expensive intervention program called the Milwaukee Project, and claimed that the program caused a boost of 30 IQ points in an experimental sample of poor black children. The remarkably strong treatment effects aroused skepticism, and there were various methodological criticisms of the study, along with criticisms of the fact that many crucial details of the study were never published. Later, Heber received a felony conviction and was imprisoned as it turned out that he and his colleagues had appropriated large amounts of the project's funding for personal use. See Page (2007) for a discussion of the Milwaukee Project.

17. Wheeler's results regarding age-related IQ decline are somewhat confounded by the Flynn effect, of which his study may be the first empirical demonstration.

18. The paucity of behavioral genetic research on blacks is a long-standing problem in the study of black-white differences, but there are now more than a handful of studies on the heritability of IQ in blacks. In the meta-analysis by Pesta et al. (2020), the mean estimates for blacks and whites in the US were as follows (K=15–16, median age=12):

Race	Heritability	Shared environment	Residual
White	.58	.20	.24
Black	.60	.15	.25

These estimates indicate that genetic and environmental effects explain very similar proportions of IQ variance in blacks and whites.

I am generally skeptical of the idea that gene-environment interactions would be an important source of variance for complex traits like IQ. This is because the genetic basis of IQ is highly polygenic, with each polymorphism contributing very little to population variance. "Genetic strains" in the sense that Jensen speaks of do not really exist for IQ. Everyone differs from

everyone else on numerous IQ loci, and no locus contributes much to population variance, so the idea that there would be “genetic strains” with different responses to the same environment is incoherent. Everyone (save identical twins) is a genetic strain unto himself.

The only situation where strong gene-environment interactions (in the sense of decreased genetic variance) can confidently be predicted are conditions of extreme deprivation where genetic variance is unlikely to be expressed, in accordance with Jensen’s threshold model. I suspect that milder deprivation is more likely to lead not to permanent deficiency in cognitive abilities but to the failure of particular measurement instruments when applied to individuals from culturally discordant backgrounds. Without tests for measurement invariance, claims of gene-environment interaction effects are difficult to interpret.

19. The twin deficit in IQ was a robust finding in older literature, but in newer studies it is either not found or is very small in magnitude, suggesting that whatever pre- and perinatal influences depressed twins’ IQs in earlier cohorts, they are generally no longer operative. See Christensen et al. (2006), Webbink et al. (2008), Calvin et al. (2009), and Eriksen et al. (2012). Another possibility is that these days twins are increasingly born to socioeconomically advantaged mothers who received IVF treatment and whose genetic IQ advantage statistically offsets the twin deficit, but within-family comparisons in the cited studies argue against that explanation.

Given that only variances and covariances and not means are included in most behavior genetic models, twin-singleton differences in means are generally not considered as particularly problematic for the generalizability of twin results.

20. The birth-order effect is observed in modern studies, too. Rohrer et al. (2015) found a 1.5 point average decrement in IQ for each increase in birth-order position. While Jensen dismissed a genetic interpretation of this effect, accumulation of harmful mutations in the gametes of older parents is in fact a plausible (partial) explanation.

21. Chmielewski (2019) provides an overview of the relation between test scores and parental socioeconomic status across the world from the 1960s up until today. She uses data from various tests of educational achievement, which are not quite IQ data but are in any case highly correlated with them (Deary et al., 2007; Kaufman et al., 2012). In the US over the last 50 years, there has been a slight diminution in the association between IQ and SES, measured as the test score gap between the 10th and 90th parental education percentiles:

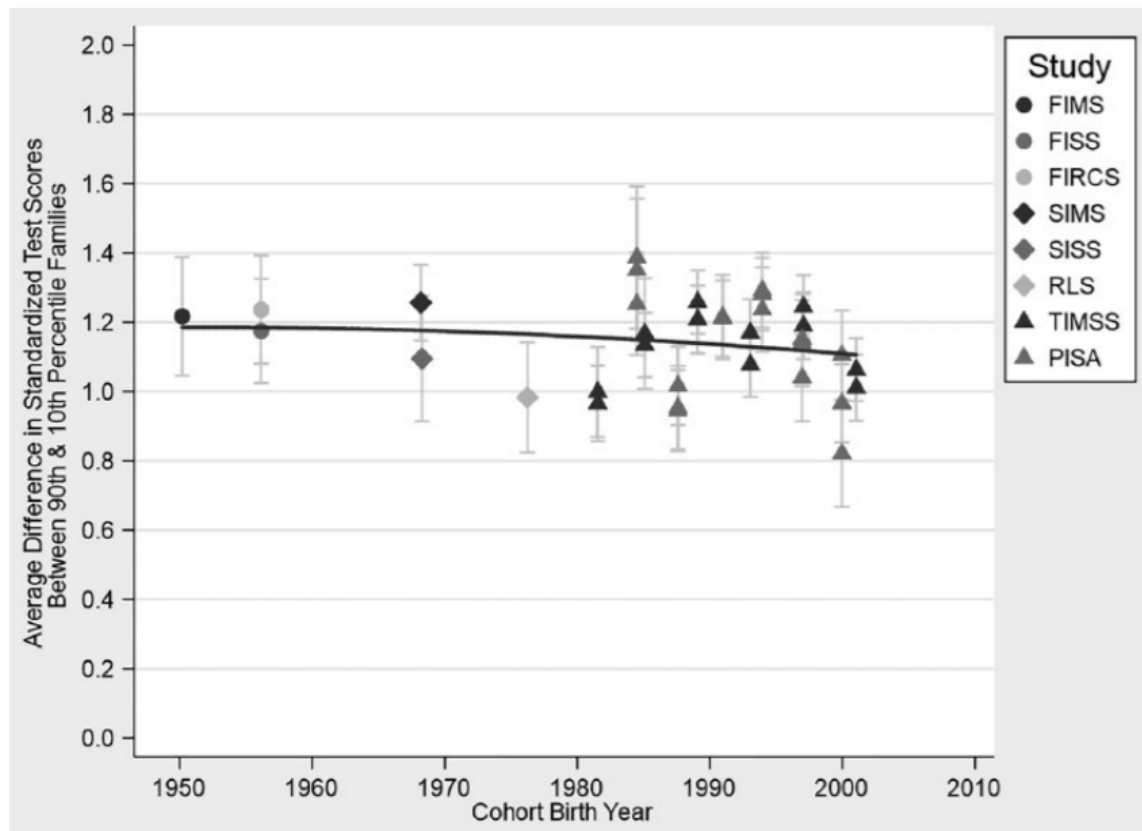


Figure 1. Trend in 90/10 Parent Education Achievement Gaps, United States, 1950 to 2001 Cohorts

Note: Gaps and quadratic fit line adjusted for age of testing and subject. Gray brackets are 95 percent confidence intervals.

Globally (N=100 countries), SES disparities in test scores have generally increased over the last few generations:

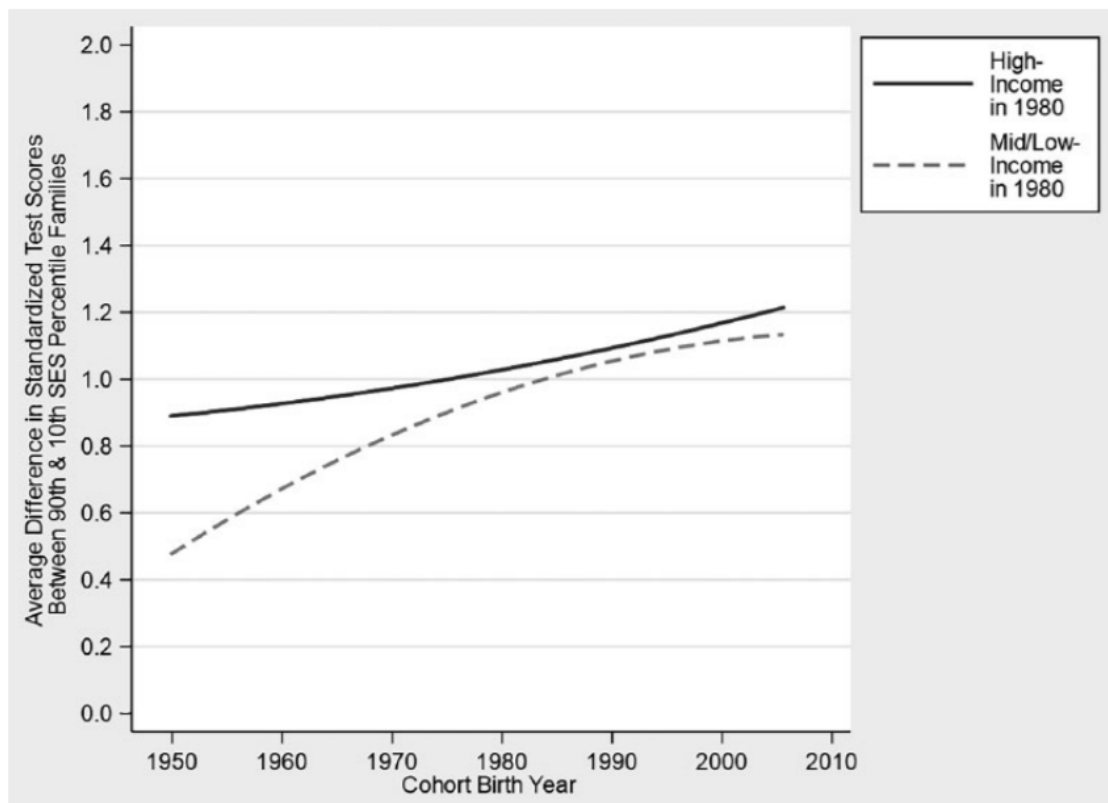


Figure 2. Estimated Quadratic Trends in 90/10 SES Achievement Gaps, by Country Income-Level, 1950 to 2005 Birth Cohorts

Note: “High-income” countries had GDPs per capita of at least \$6,000 in 1980 (see Table A1 in the online supplement for coding). Trend lines are estimates from Model 5 (Table 2). Fixed values for control variables: SES = parent education, subject = math, all others = 0 or reference category.

Chmielewski links these growing disparities to “rapidly increasing school enrollments” which reveal “educational inequality that was previously hidden outside the school system.” What this shows is that the basic problem of cognitive inequality which was at the center of Jensen’s

intellectual project shows no signs of abating anywhere in the world.

The most audacious estimates for the permanence of socioeconomic and hence also cognitive stratification come from Greg Clark (2014). Using surname data, he claims that the intergenerational elasticity of social status is around .75 across many countries and long historical periods, from medieval England to modern Japan. Whereas conventional estimates of social mobility suggest that social classes in modern societies are in a state of constant churn, with most family lineages rising and falling and rising again in terms of status every few generations, surname estimates imply that families usually retain their status for centuries. According to Clark, this stability is masked by the fact that any single indicator of status, such as income or occupation, is very imperfectly correlated with the underlying social competence variable. Clark argues that the most parsimonious explanation of these findings is that the underlying variable has high additive heritability. However, for lineages to “breed true” to the extent that Clark finds, assortative mating with respect to the underlying genetic component must be strong.

22. The key message of the Coleman Report, which was based on data from more than 600,000 students, was that family background was the major determinant of academic achievement, and that differences between schools had little effect independently of the students’ own characteristics. This conclusion contradicted not only Coleman and colleagues’ own theoretical assumptions, but also those of social science in general, and it was a blow to the “Great Society” ideology, the dream of eliminating inequality through social engineering.

Morgan and Jung (2016) provide an update on the Coleman report. Using data from the Education Longitudinal Study that began in 2002, they affirm that differences between schools in available resources continue to have little effect on educational achievement and attainment, and that family background remains the primary determinant of student outcomes.

23. The failure rates of blacks and whites on the AFQT imply a standardized gap of $d=1.35$. See [here](#) for a simple Excel formula for making this calculation.

24. Fertility appears to have been more dysgenic in blacks than whites for some time in America, but it does not seem to have expanded the black-white IQ gap in recent cohorts. This is probably because of various countervailing influences, such as improved maternal and perinatal health care, increasing admixture with whites, and cognitively selective immigration of blacks from the West Indies and Africa.

25. According to the latent variable perspective prevalent in psychometrics, IQ test items are not of any special interest by themselves but rather because they tap into latent abilities that cannot be directly measured. Teaching children how to answer specific items therefore dilutes the validity of the test, similarly to how the use of a cheat sheet will invalidate an exam as an indicator of what the student has learned. Spitz (1986, p. 146) recounted how “teaching to the test” reached its *reductio ad absurdum* in intervention programs devised by behaviorists:

John Throne and his colleagues at the University of Kansas have added some bewildering twists to the empiricism of Skinner and Bijou. Not only can intelligent behavior be produced using operant conditioning (selective reinforcement), but mental retardation—defined as a behavioral disorder—can be reversed by behavioral training [...]. Because performance on intelligence tests is the accepted measure, and because from a behavioral perspective intelligent performance is identical with intelligence as an abstraction, the way to reverse mental retardation is to train retarded children to perform well on the 12 subtests of the Wechsler Intelligence Intelligence Scale for Children!

26. te Nijenhuis et al. (2014) found a negative correlation between *g* loadings and test performance gains across eight Head Start studies, suggesting that compensatory education programs tend to boost specific skills rather than *g*. This finding is in line with many other studies showing that education and training improve the specific skills targeted, rather than leading to generalized ability gains (Ritchie et al., 2015; Sala & Gobet, 2017; Dick et al., 2019). The significance of high *g* is that it gives the individual an edge across diverse domains without any prior training, while also making skill and knowledge acquisition easier.

27. Protzko's (2015) meta-analysis of 39 randomized controlled trials confirms the fadeout effect. The experimental group gets an IQ boost which gradually dissipates after the experiment ends, and the IQs of the experimental group converge with those of the control group.

28. Spitz (1999) offers a dyspeptic post-mortem of the long-lived Pygmalion controversy.

29. Overall, Jensen's 1969 article has stood the test of time quite well. The Level I/Level II ability distinction is, however, undoubtedly outdated, and the emphasis Jensen put on it was misplaced. The distinction is not a useful way of thinking about race and SES differences in cognitive ability. Jensen himself later reflected on the development of his thinking on Level I and II abilities in this way:

Level I/Level II was important, I think, because it revealed a type of mental ability in which black-white differences are minimal as compared with the ability (or abilities) measured by traditional IQ tests or similar highly *g*-loaded cognitive tasks. It suggested what then seemed a promising possibility, that the Level I/Level II distinction might lend itself to an aptitude-by-training interaction that could decrease the disparity in scholastic performance between typical black and white children. This hope has not panned out, I conjecture, because of the intrinsically highly *g*-loaded, or Level II, nature of educational achievement. Educational achievements seem to be valued almost directly to the extent that they are perceived as *g*-loaded, and this is as true for any ethnic minorities in our schools as for the white majority.

In recent years I have placed less emphasis on the Levels hypothesis, which I now view as merely a special case of what I regard as a much broader and more fundamental phenomenon that Spearman first noted in 1927 and which I have termed ‘Spearman’s hypothesis’. This hypothesis states that the black-white difference is essentially a difference in *g*, and the varying magnitudes of the mean black-white differences (in standard-score units) on various tests are directly related to the tests’ *g* loadings. A preponderance of evidence substantiates Spearman’s hypothesis, although there is also evidence that certain other factors independent of *g*, such as spatial visualization ability, also show mean black-white differences, but to a much lesser degree than the *g* difference [...]. Hence Level II can be equated with Spearman’s *g* and Level I represents only a fairly narrow category of tasks (rote learning and memory span) among all those tasks that show especially low loadings on *g*. I still think it worthwhile to investigate the broad realm of very low *g*-loaded cognitive tasks in relation to various population differences, with a view to discovering abilities that may afford some educational and occupational leverage for individuals who fall markedly below the norm of performance on highly *g* loaded tasks.

–Jensen (1987)

30. The situation is not necessarily that much better these days. Warne et al. (2018, 2019) show that courses dedicated to intelligence are rare in American colleges and universities and that the coverage of the topic in psychology textbooks is often inaccurate. See also Hunt (2013).

31. After *The g Factor*, Jensen published one more book, the 2006 *Clocking the Mind: Mental Chronometry and Individual Differences*. It presented his final attempt to anchor IQ research in lower-level cognitive constructs. I think this endeavor, motivated in part by Jensen’s long-term desire to develop a [ratio scale](#) of intelligence, was unsuccessful for the same reasons that cognitive psychology has generally failed to devise credible models of complex cognitive functions: even the simplest cognitive tasks that a researcher may employ are imperfect reflections of latent variables rather than isomorphic with them. Deary (2000) and Deary et al. (2016) provide a cogent refutation of this sort of “greedy reductionism” that has hampered cognitive psychology.

32. It would be feasible to test the extent of genetic selection that immigration has imposed on different populations, given that currently available polygenic scores for IQ have substantial validity in non-whites, too. You could, say, take a random sample of Nigerian immigrants to America, obtain polygenic scores from them, and then compare those to scores obtained from a random sample of people in Nigeria. The difference between the samples, corrected for unreliability, would give an estimate of genetic selection.

33. The admissions test data are from [here](#). The d values were computed using [this formula](#), which assumes normality within groups and equal variances across groups.

34. *Herrnstein's syllogism* encapsulates the logic of hereditarianism with respect to SES differences:

- 1) If differences in mental abilities are inherited, and
- 2) If success requires those abilities, and
- 3) If earnings and prestige depend on success,
- 4) Then social standing (which reflects earnings and prestige) will be based to some extent on inherited differences among people.

—Herrnstein et al. (1973)

35. I wrote about regression to the mean in the context of quantitative genetic models [here](#).

36. While blacks are naturally disproportionately represented in the lower end of the IQ distribution, too, there is an interesting racial wrinkle associated with low mental ability. As shown by Paul Morgan and colleagues in several large studies (e.g., Morgan et al., 2017), white children are more likely to be identified as intellectually or otherwise disabled than non-white children matched for test scores and various background variables. This may be partly because white parents are more proactive about securing special services for their poorly-achieving children. Another reason may be that school authorities are leery of labeling non-whites as disabled. There has long been concern about black children in particular being overidentified as mentally disabled, with various policies enacted to address the problem. This may have led to an opposite problem of underidentification.

37. Some critics have pointed to the small amount of variance in IQ—0.7 percent in Model 1—that is explained by white ancestry in African-Americans, suggesting that this shows something to be amiss in the analysis by Lasker et al. This criticism is based on a misunderstanding of the logic of admixture analysis. As Lasker et al. explain, an R^2 of 0.7 percent, which corresponds to a correlation of 0.086 between IQ and white ancestry, is entirely compatible with most of the IQ gap being genetic. The difference in white ancestry between blacks and whites in the sample is 6.83 standard deviations, in units of variation in white ancestry in the black sample. $0.086 \times 6.83 \approx 0.59$ standard deviations, which means that around 60 percent of the one standard deviation IQ gap can be attributed to genetic differences.

The hereditarian model does not claim that white ancestry is a major source of IQ variation among African-Americans. Rather, white admixture is expected to cause a minor upward shift in

IQ in African-Americans—whose IQ distribution is mostly determined by the distribution of the IQ alleles they have inherited from their African ancestors—and this shift can be used to estimate the heritability of the IQ gap.

A criticism of the admixture method that is sometimes proffered is that because we do not know how representative in terms of IQ alleles the white ancestors of African-Americans are, the white ancestry component in African-Americans cannot be used as a test of the hereditarian model. However, for the genetic composition of those white ancestors (who were mostly men from the period of black slavery) to seriously bias the test in favor of the hereditarian model, only the intellectual elite of the South would have had to have opportunities for mating with blacks, which, Thomas Jefferson's alleged proclivities notwithstanding, strikes me as very unlikely. If, on the other hand, the white ancestors were below-average in IQ, the admixture test provides a conservative estimate of the genetic effect.

It seems that it should be possible, at least in principle, to test if the IQ allele distribution of the white ancestors of American blacks differs from that of whites in the general population. The method would involve identifying stretches of DNA in blacks that are inherited from whites, and comparing if the average polygenic score computed from IQ-associated alleles that happen to fall within those stretches is significantly different from the white average. I am not sure if this method is feasible in practice, though.

38. Turkheimer's laws are:

? First Law. All human behavioral traits are heritable.

? Second Law. The effect of being raised in the same family is smaller than the effect of genes.

? Third Law. A substantial portion of the variation in complex human behavioral traits is not accounted for by the effects of genes or families.

39. Turkheimer's views on the biometric decomposition of group differences are somewhat inconsistent across different publications. In Turkheimer (1991), he endorsed Jensen's monistic view of group and individual differences in the context of adoption studies. In his most cited paper (Turkheimer et al., 2003) and follow-up papers to it, he treats the moderation of biometric parameters by a group-level variable (SES) as feasible. Yet elsewhere, as in the 2019 blog post discussed above, he dismisses the biometric analysis of group differences as unfeasible. It is difficult to not suspect that his incoherence is influenced by [extrascientific concerns](#).

40. The statement was published as a paid advertisement on September 28, 1973, and can be read [here](#). The list of signatories is available [here](#).

41. While different selective pressures acting on the ancestors of Europeans and West Africans especially in prehistoric times are by far the most plausible ultimate explanation under the hereditarian model, many other scenarios are entirely compatible with hereditarianism. For example, both whites and Africans who moved, voluntarily or forcibly, to America may have been non-representative of their populations origin, causing genotypic IQ differences in their descendants. Selective pressures in America during the last few centuries may also have played a role. Even if we assumed that genotypic IQ has only been affected by genetic drift since the divergence of different human lineages, hereditarianism with respect to race [would still be viable](#).

42. Some neuroscience research on IQ is admittedly of good quality, namely that conducted by differential psychologists—see e.g., Cox et al. (2019). Nevertheless, even at its best, neuroscience is complementary to psychology, not something that could replace it.

43. The larger than usual black-white gaps found in many affluent, liberal cities are surely mostly a function of parental ability selection and differential racial regression to the mean. However, it is additionally possible that progressive educational philosophies often favored in such places emphasize student independence and discourage strict discipline to the particular detriment of black students.

44. Reardon et al. do not argue against hereditarianism but simply point to Nisbett et al. (2012) as the justification for their blank slate perspective. The paper by Nisbett et al. is a supposed update of the 1996 American Psychological Association Task Force report on intelligence (Neisser et al., 1996). The 1996 report is still a worthwhile overview of intelligence research thanks to the balanced mix of genetically- and environmentally-oriented researchers among its authors. Appropriately, the APA report gave firm answers where the evidence base was (and is) strong and was more equivocal where there was less evidence.

In contrast, the 2012 paper by Nisbett et al. is already largely outdated. It is a compendium of the authors' pet anti-hereditarian arguments. It uses findings from small, often peculiarly old and generally unreplicated studies in an attempt to undermine some of the firmest conclusions in IQ research. For example, the paper claims that the black-white IQ gap has narrowed considerably, something that does not find support in recent data. It also puts stock in stereotype threat as an explanation of the black-white gap, but that idea is already on its last legs (e.g., Shewach et al., 2019) and will likely receive a *coup de grâce* in the [upcoming, preregistered replication study](#).

The general philosophical problem in the Nisbett et al. paper is that rather than presenting a comprehensive model of IQ, including its relation to race, it ignores the big picture and lists lots of arguments of generally low quality in hopes that this concatenation of unconnected claims will by sheer force of its volume cohere into a refutation of hereditarianism. James Lee did a masterful fisting of this style of anti-hereditarianism in his [review](#) of Nisbett's book *Intelligence and How to Get It* (Lee, 2009).

The way Reardon et al. describe the state of the hereditarian versus environmental debate demonstrates either gross scholarly incompetence or strategic dishonesty.

45. Another method for studying the genetic basis of international IQ differences is that pioneered by Davide Piffer (e.g., Piffer, 2019). His approach is to regress national IQs on average national polygenic scores for educational attainment, using IQ data from Richard Lynn and colleagues, and DNA data from public repositories such as the Human Genome Diversity Project. His analyses show that there is a very high correlation between national IQ and national mean polygenic scores, the latter computed based on genome-wide association study results from European samples.

An obvious criticism of Piffer's approach is that because the available polygenic scores are based on European discovery samples, they are biased measures of underlying polygenic propensities in non-Europeans. It is, after all, well-known that the predictive validity of polygenic scores decays when you use them in populations ancestrally distant from the discovery population. Lasker et al. (2019), for example, found that the correlations between polygenic scores (based on allelic effect sizes from white discovery samples) and IQ were around 0.04–0.11 and 0.21–0.23 in African-Americans and European-Americans, respectively.

The leading explanation of why the predictive power of polygenic scores is not invariant across populations is that many phenotype-associated polymorphisms discovered in GWAS research are not causal but rather are only correlated with the true, causal polymorphisms. This is because genetic variants are not inherited singly but in [haplotype blocks](#), making it challenging to pinpoint true causal variants. Therefore, much of the validity of polygenic scores is due to them tapping into useful *proxies* of causal variants. However, when polygenic scores are used in a population that is ancestrally dissimilar to the discovery population, with a completely different correlation structure between alleles, there is no reason to expect the same variants to be useful as proxies for causal variants, leading to decay in predictive power.

We are in the early days of polygenic scoring, and various biases in the scores are not well understood. Therefore, I do not think polygenic scoring is among the strongest methods in the race realist armamentarium at the moment. Nevertheless, there is a mathematical argument that could imply that Piffer's results are approximately valid despite the problematic between-race generalizability of polygenic scores. This would be the case if the error introduced to the scores by the proxy-variant problem is *random* in nature. In that case, the frequencies of some causal variants will be overestimated in non-Europeans, while the frequencies of some others will be underestimated, with the result that the mean values are estimated more or less correctly. This is, of course, a basic property of the linear regression model: random error in the dependent variable increases the variance of the estimate but it does not bias it.

References

Ang, S., Rodgers, J. L., & Wänström, L. (2010). The Flynn effect within subgroups in the U.S.: Gender, race, income, education, and urbanization differences in the NLSY-Children data. *Intelligence*, 38, 367–384.

Barnes, J. C., Wright, J. P., Boutwell, B. B., Schwartz, J. A., Connolly, E. J., Nedelec, J. L., & Beaver, K. M. (2014). Demonstrating the validity of twin research in criminology. *Criminology*, 52, 588–626.

Beaujean, A. A., & Osterlind, S. J. (2008). Using item response theory to assess the Flynn effect in the National Longitudinal Study of Youth 79 children and young adults data. *Intelligence*, 36, 455–463.

Beaujean, A. A., & Sheng, Y. (2014). Assessing the Flynn effect in the Wechsler scales. *Journal of Individual Differences*, 35, 63–78.

Beaver, K.M., Schwartz, J.A., Connolly, E.J., Nedelec, J.L., Al-Ghamdi, M.S., & Kobeisy, A.N. (2013). The genetic and environmental architecture to the stability of IQ: Results from two independent samples of kinship pairs. *Intelligence*, 41, 428–438.

Bouchard, T. J., Jr. (1997). IQ similarity in twins reared apart: Findings and responses to critics. In R. J. Sternberg & E. L. Grigorenko (Eds.), *Intelligence, heredity, and environment* (pp. 126–160). Cambridge, England: Cambridge University Press.

Branigan, A. R., McCallum, K. J., & Freese, J. (2013). Variation in the heritability of educational attainment: An international meta-analysis. *Social Forces*, 92, 109–140.

Brown, T. A. (2014). *Confirmatory Factor Analysis for Applied Research*. New York, NY: Guilford Publications.

Burks, B. S. (1928). The relative influence of nature and nurture upon mental development: A comparative study of foster parent-foster child resemblance and true parent-true child resemblance. *Yearb. Nat. Soc. Stud. Educ.*, 27, 219–316.

Button, K. S. et al. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. *Nat. Rev. Neurosci.*, 14, 365–376.

Calvin, C., Fernandes, C., Smith, P., Visscher, P. M., & Deary, I. J. (2009). Is there still a cognitive cost of being a twin in the UK? *Intelligence*, 37(3), 243–248.

Cesarini, D. (2010). *Essays on Genetic Variation and Economic Behavior*. Massachusetts Institute of Technology.

Chmielewski, A. K. (2019). The Global Increase in the Socioeconomic Achievement Gap, 1964 to 2015. *American Sociological Review*, 84, 517–544.

Christensen, K., Petersen, I., Skytthe, A., Herskind, A. M., McGue, M., & Bingley, P. (2006). Comparison of academic performance of twins and singletons in adolescence: follow-up study. *Bmj*, 333(7578), 1095.

Clark, G. (2014). *The Son Also Rises: Surnames and the History of Social Mobility*. Princeton, NJ: Princeton University Press.

Conley, D., Rauscher, E., Dawes, C., Magnusson, P. K., & Siegal, M. L. (2013). Heritability and the equal environments assumption: evidence from multiple samples of misclassified twins. *Behavior Genetics*, 43, 415–426.

Cox, S. R. et al. (2019). Structural brain imaging correlates of general intelligence in UK Biobank. *Intelligence*. 76, 101376.

Cronbach, L. J., & Meehl, P. C. (1955). Construct validity in psychological tests. *Psychological Bulletin*, 52, 281–302.

de Zeeuw, E. L., de Geus, E. J., & Boomsma, D. I. (2015). Meta-analysis of twin studies highlights the importance of genetic variation in primary school educational achievement. *Trends in Neuroscience and Education*, 4, 69–76.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. New York: Oxford University Press.

Deary I. J., Cox S. R., & Ritchie S. J. (2016) Getting Spearman off the Skyhook: One More in a Century (Since Thomson, 1916) of Attempts to Vanquish *g*. *Psychol. Inq.*, 27, 192–199.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence*, 35, 13–21.

Derks, E. M. et al. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics*, 9, 403–411.

Dick, A. S., et al. (2019). No evidence for a bilingual executive function advantage in the nationally representative ABCD study. *Nat. Hum. Behav.*, 3, 692.

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

Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, 35, 21–50.

Dolan, C. V., & Hamaker, E. L. (2001). Investigating Black–White differences in psychometric IQ: Multi-group confirmatory factor analyses of the WISC-R and K-ABC and a critique of the method of correlated vectors. In Frank Columbus (Ed.), *Advances in Psychology Research*, vol. 6 (pp. 31–59). Huntington, NY: Nova Science Publishers.

Dolan, C. V., Huijskens, R. C., Minica, C. C., Neale, M. C., & Boomsma, D. I. (2019). Incorporating polygenic scores in the twin model to estimate genotype-environment covariance: exploration of statistical power. *bioRxiv*, 702738.

Dolan, C. V., Molenaar, P. C. M., & Boomsma, D. I. (1992). Decomposition of multivariate phenotypic means in multigroup genetic covariance structure analysis. *Behavior Genetics*, 22, 319–335.

Dreger, R. M., & Miller, K. S. (1960). Comparative psychological studies of Negroes and whites in the United States. *Psychol. Bull.*, 57, 361-402.

Dreger, R. M., & Miller, K. S. (1968). Comparative psychological studies of Negroes and whites in the United States: 1959-1965. *Psychol. Bull. (Monogr. Suppl. 70, No. 3, Part 2)*.

Engelhardt, L. E., Mann, F. D., Briley, D. A., Church, J. A., Harden, K. P., & Tucker-Drob, E. M. (2016). Strong genetic overlap between executive functions and intelligence. *Journal of Experimental Psychology: General*, 145, 1141–1159.

Eriksen, W., Sundet, J. M., & Tambs, K. (2012). Twin–Singleton Differences in Intelligence: A Register-Based Birth Cohort Study of Norwegian Males. *Twin Research and Human Genetics*, 15, 649–655.

Erlenmeyer-Kimling, L., & Jarvik, L. F. (1963). Genetics and intelligence: a review. *Science*, 142, 1477–1479.

Felson, J. (2014). What can we learn from twin studies? A comprehensive evaluation of the equal environments assumption. *Social Science Research*, 43, 184–99.

Fox, M. C., & Mitchum, A. L. (2013). A knowledge-based theory of rising scores on “culture-free” tests. *Journal of Experimental Psychology: General*, 142, 979–1000.

Fox, M. C., & Mitchum, A. L. (2014) Confirming the Cognition of Rising Scores: Fox and Mitchum (2013) Predicts Violations of Measurement Invariance in Series Completion between Age-Matched Cohorts. *PLOS ONE* 9(5): e95780.

Franic, S., et al. (2013). Can genetics help psychometrics? Improving dimensionality assessment through genetic factor modeling. *Psychological Methods*, 18, 406–433.

Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS-IV/WMS-IV standardization data. *Intelligence*, 51, 79–97.

Fuerst, J. (2015). *Nature of Race*. Open Behavior Genetics.

Funder, D. C., & Ozer, D. J. (2019). Evaluating effect size in psychological research: Sense and nonsense. *Advances in Methods and Practices in Psychological Science*, 2, 156–168.

Giessman, J. A., Gambrell, J. L., & Stebbins, M. S. (2013). Minority performance on the Naglieri Nonverbal Ability Test, Second Edition, versus the Cognitive Abilities Test, Form 6: One gifted program's experience. *Gifted Child Quarterly*, 57, 101–1009.

Gottschling, J., Hahn, E., Beam, C. R., Spinath, F. M., Carroll, S., & Turkheimer, E. (2019). Socioeconomic status amplifies genetic effects in middle childhood in a large German twin sample. *Intelligence*, 72, 20–27.

Herrnstein, R. J., Deutsch, K. W., & Edsall, T. B. (1973). Controversy. *Society*, 10, 5–6.

Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. NY: Free Press.

Hill, A. B. (1965). The environment and disease: association or causation? *Proc R Soc Med*, 58, 295–300.

Hill, W. D., et al. (2018). Genomic analysis of family data reveals additional genetic effects on intelligence and personality. *Molecular Psychiatry*, 23, 2347–2362.

Hunt, E. (2014). Teaching intelligence: Why, why it is hard and perhaps how to do it. *Intelligence*, 42, 156–165.

Hur, Y., & Bates, T. (2019). Genetic and Environmental Influences on Cognitive Abilities in Extreme Poverty. *Twin Research and Human Genetics*, 22, 297–301.

Jensen, A. R. (1958). Personality. *Annual Review of Psychology*, 9, 295–322.

Jensen, A. R. (1969a). How Much Can We Boost IQ and Scholastic Achievement? *Harvard Educational Review*, 39, 1–123.

Jensen, A. R. (1969b). Reducing the heredity-environment uncertainty. *Harvard Educational*

Review, 39, 449–483.

Jensen, A. R. (1973). On “Jensenism”: A reply to critics. In B. Johnson (Ed.), *Education yearbook, 1973-74*. New York: Macmillan Educational Corporation. Pp. 276–298.

Jensen, A. R. (1978). Citation Classics (How much can we boost IQ and scholastic achievement?). *Current Contents*, 41, 16.

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

Jensen, A. R. (1985). The nature of the black-white difference on various psychometric tests: Spearman’s hypothesis. *Behavioral and Brain Sciences*, 8, 193–219.

Jensen, A. R. (1987). Differential psychology: Towards consensus. In M. Modgil and C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy*. London: Falmer Press.

Jensen, A. R. (1998a). Jensen on “Jensenism”. *Intelligence*, 26, 181–208.

Jensen, A. R. (1998b). *The g factor: The science of mental ability*. Westport, CT: Praeger.

Jensen, A. R. (2006). *Clocking the Mind: Mental Chronometry and Individual Differences*. Amsterdam: Elsevier.

Johnson, W. (2012). How much can we boost IQ? An updated look at Jensen’s (1969) question and answer. In *Developmental psychology: Revisiting the classic studies*. SAGE, London, 118–131.

Kaufman, S.B., Reynolds, M.R., Liu, X., Kaufman, A.S., & McGrew, K.S. (2012). Are cognitive g and academic achievement g one and the same g? An exploration on the Woodcock-Johnson and Kaufman tests. *Intelligence*, 40, 123–138.

Keller, M. C., Garver-Apgar, C. E., Wright, M. J., Martin, N. G., Corley, R. P., Stallings, M. C., Hewitt, J. K., & Zietsch, B. P. (2013) The genetic correlation between height and IQ: Shared genes or assortative mating? *PLOS Genetics*, 9, 1–10.

Kirkegaard, E. O. W. (2019). Race Differences: A Very Brief Review. *The Mankind Quarterly*, 60, 142–173.

Kirkegaard, E. O. W., Wang, M., & Fuerst, J. (2018). Biogeographic Ancestry and Socioeconomic Outcomes in the Americas: A Meta-Analysis. *Mank. Q.*, 57, 398–427.

Krapohl, E. et al. (2014). The high heritability of educational achievement reflects many

genetically influenced traits, not just intelligence. *Proc Natl Acad Sci*, 111, 15273–15278.

Lasker, J. et al. (2019). Global Ancestry and Cognitive Ability. *Psych*, 1, 431–459.

Lee, J. J. (2009). Review of intelligence and how to get it: Why schools and cultures count. *Personality and Individual Differences*, 48, 247–255.

Lee, T., Mosing, M.A., Henry, J.D. et al. (2012). Genetic Influences on Five Measures of Processing Speed and Their Covariation with General Cognitive Ability in the Elderly: The Older Australian Twins Study. *Behav Genet*, 42, 96–106.

Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences and measurement invariance in the context of the common factor model. *Intelligence*, 31, 543–566.

Mazumder, B. (2014). Black-White Differences in Intergenerational Economic Mobility in the US. *Economic Perspectives*, 38, 1–18.

McGue, M., Rustichini, A., Iacono W. G. (2017). Cognitive, noncognitive, and family background contributions to college attainment: a behavioral genetic perspective. *J Pers*, 85, 65–78.

Mollon, J., Knowles, E.E.M., Mathias, S.R. et al. (2018). Genetic influence on cognitive development between childhood and adulthood. *Mol Psychiatry*.

Morgan, P. L., et al. (2017). Replicated Evidence of Racial and Ethnic Disparities in Disability Identification in U.S. Schools. *Educational Researcher*, 46, 305-322.

Morgan, S. L. & Jung, S. B. (2016). Still No Effect of Resources, Even in the New Gilded Age?. *Russell Sage Foundation Journal of the Social Sciences*, 2, 83-116.

Murray, C. (1999). The Secular Increase in IQ and Longitudinal Changes in the Magnitude of the Black-White Difference: Evidence from the NLSY. Paper presented to the Behavior Genetics Association Meeting, 1999.

Murray, C. (2007). The Magnitude and Components of Change in the Black-White IQ Difference from 1920 to 1991: A Birth Cohort Analysis of the Woodcock-Johnson Standardizations. *Intelligence*, 35, 305–18.

Must, O., te Nijenhuis, J., Must, A., & van Vianen, A. E. M. (2009). Comparability of IQ scores over time. *Intelligence*, 37, 25–33.

te Nijenhuis, J., Jongeneel-Grimen, B., & Kirkegaard, E.O. (2014). Are headstart gains on the g factor? A meta-analysis. *Intelligence*, 46, 209–215.

Page, E. B. (2007). Milwaukee Project. In C. R. Reynolds & E. Fletcher-Janzen (Eds.), *Concise Encyclopedia of Special Education: A Reference for the Education of the Handicapped and Other Exceptional Children and Adults* (pp. 1366–1368). Hoboken, NJ: Wiley.

Panizzon, M. S., et al. (2014). Genetic and environmental influences on general cognitive ability: Is g a valid latent construct? *Intelligence*, 43, 65–76.

Pesta, B., et al. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence*, 78, p. xx–xx.

Pietschnig, J., Tran, U. S., & Voracek, M. (2013). Item-response theory modeling of IQ gains (the Flynn effect) on crystallized intelligence: Rodgers' hypothesis yes, Brand's hypothesis perhaps. *Intelligence*, 41, 791–801.

Piffer, D. (2019). Evidence for Recent Polygenic Selection on Educational Attainment and Intelligence Inferred from Gwas Hits: A Replication of Previous Findings Using Recent Data. *Psych*, 1, 55–75.

Polderman, T. J. C., et al. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies *Nat. Genet.*, 47, 702–709.

Protzko, J. (2015). The environment in raising early intelligence: A meta-analysis of the fadeout effect. *Intelligence*, 53, 202–210.

Rinderman et al. (2020). Survey of expert opinion on intelligence: Intelligence research, experts' background, controversial issues, and the media. *Intelligence*, 78, p. xx–xx.

Rohrer, J. M., Egloff, B., & Schmukle, S. C. (2015). Examining the effects of birth order on personality. *Proceedings of the National Academy of Sciences*, 112, 14224–14229.

Rowe, D. C., & Cleveland, H. H. (1996). Academic achievement in Blacks and Whites: Are the developmental processes similar? *Intelligence*, 23, 205–228.

Sala, G., & Gobet, F. (2017). Does far transfer exist? Negative evidence from chess, music and working memory training. *Current Directions in Psychological Science*, 26, 515–520.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European community. *Journal of Applied Psychology*, 88, 1068–1081.

- Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, 86, 162–173.
- Segestråle, U. (2000). *Defenders of the truth. The sociobiology debate*. New York: Oxford University Press.
- Sesardic, N. (2005). *Making sense of heritability*. Cambridge, UK: Cambridge University Press.
- Shewach, O. R., Sackett, P. R., & Quint, S. (2019). Stereotype threat effects in settings with features likely vs. unlikely in operational testing settings: A meta-analysis. *Journal of Applied Psychology*, 104, 1514–1534.
- Shiu, W., et al. (2013). An item-level examination of the Flynn effect on the National Intelligence Test in Estonia. *Intelligence*, 41, 770–779.
- Shuey, A. M. (1966). *The testing of Negro intelligence*. (2nd ed.) New York: Social Science Press.
- Silventoinen, K., Posthuma, D., van Beijsterveldt, T., Bartels, M., & Boomsma, D. I. (2006). Genetic contributions to the association between height and intelligence: Evidence from Dutch twin data from childhood to middle age. *Genes, Brain, & Behavior*, 5, 585–595.
- Skeels, H. M., & Dye, H. B. (1939). A study of the effects of differential stimulation on mentally retarded children. *Proceedings and Addresses of the American Association on Mental Deficiency*, 44, 114–136.
- Snyderman, M., & Rothman, S. (1987). Survey of expert opinion on intelligence and aptitude testing. *American Psychologist*, 42, 137–144.
- Snyderman, M., & Rothman, S. (1990). *The IQ Controversy, the Media, and Public Policy*. New Jersey: Transaction Publishers.
- Spitz, H. H. (1986). *The raising of intelligence: A selected history of attempts to raise retarded intelligence*. Hillsdale, NJ: Erlbaum.
- Spitz, H. H. (1999). *Beleaguered Pygmalion: A History of the Controversy Over Claims that Teacher Expectancy Raises Intelligence*. *Intelligence*, 27, 199–234.
- Trundt, K. M. (2013). *Construct Bias in the Differential Ability Scales, Second Edition (DAS-II): A comparison among African American, Asian, Hispanic, and White Ethnic Groups*. Unpublished doctoral dissertation, University of Texas, Austin, TX.

Tucker-Drob E. M., Briley D. A., Harden K. P. (2013). Genetic and environmental influences on cognition across development and context. *Current Directions in Psychological Science*, 22, 349–355.

Turkheimer, E. (1991). Individual and group differences in adoption studies of IQ. *Psychological Bulletin*, 110, 392–405.

Turkheimer, E. (1997). The search for a psychometric left. *Current Psychology of Cognition*, 16, 779–783.

Turkheimer, E., Beam, C. R., Sundet, J. M., & Tambs, K. (2017). Interaction between parental education and twin correlations for cognitive ability in a Norwegian conscript sample. *Behavior Genetics*, 47, 507–515.

Turkheimer, E., Haley, A., Waldron, M., D’Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, 14, 623–628.

Vinkhuyzen, A.A., van der Sluis, S., Maes, H.H.M. et al. (2012). Reconsidering the Heritability of Intelligence in Adulthood: Taking Assortative Mating and Cultural Transmission into Account. *Behav Genet*, 42, 187–198.

Wai, J., & Putallaz, M. (2011). The Flynn effect puzzle: A 30-year examination from the right tail of the ability distribution provides some missing pieces. *Intelligence*, 39, 443–455.

Warne, R. T., Astle, M., & Hill, J. C. (2018). What do undergraduates learn about human intelligence? An analysis of introductory psychology textbooks. *Archives of Scientific Psychology*, 6, 32–50.

Warne, R. T., & Burton, J. Z. (2019). The Neglected Intelligence Course: Needs and Suggested Solutions.

Warne, R. T., Godwin, L. R., & Smith, K. V. (2013). Are there more gifted people than would be expected in a normal distribution? An investigation of the overabundance hypothesis. *Journal of Advanced Academics*, 24, 224–241.

Webbink, D., Posthuma, D., Boomsma, D. I., Geus, E. J. D., & Visscher, P. M. (2008). Do twins have lower cognitive ability than singletons? *Intelligence*, 36, 539–547.

Wheeler, L. R. A comparative study of the intelligence of East Tennessee mountain children. *J. educ. Psychol.*, 1942, 33, 321–334.

Wicherts, J., Dolan, C., Hessen, D., Oosterveld, P., van Baal, C., Boomsma, D., & Span, M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, 32, 509–537.

Wright, S. (1931) Statistical methods in biology. *J. Amer. stat. Ass.*, 26, 155–163.

Wu, A. D., Li, Z., & Zumbo, B. D. (2007). Decoding the meaning of factorial invariance and updating the practice of multi-group confirmatory factor analysis: A demonstration with TIMSS data. *Practical Assessment, Research and Evaluation*, 12, 1–26.

Share this:

- [Twitter](#)
- [Facebook](#)
- [Reddit](#)
-

[Customize buttons](#)

Human Varieties

PDF generated April 10, 2021 at 10:08 AM by Kalin's PDF Creation Station WordPress plugin