



Book Review

Review of intelligence and how to get it: *Why schools and cultures count*, R.E. Nisbett, Norton, New York, NY (2009). ISBN: 9780393065053

Abstract Richard Nisbett's *intelligence and how to get it* advances several interlocking claims: (1) the heritability of IQ is far lower than typically claimed by behavioral geneticists, (2) the IQ differences across social classes are largely environmental in origin, (3) the IQ differences across racial groups are *entirely* environmental in origin, and (4) these group differences can be narrowed substantially by interventions that social scientists have already discovered. In this review I show that Nisbett's arguments are consistently overstated or unsound.

Keywords: Intelligence; Genetics; Malleability; Nisbett

1. Heritability and mutability

A reader sympathetic to Nisbett's aims might complain that I devote excessive criticism to points on which Nisbett spends only a few lines. Nisbett's telegraphic treatment of these points, however, is precisely the glaring fault of his account. He repeatedly reaches cursory conclusions regarding the heritability of IQ that are at odds with a more searching analysis.

1.1. Twin studies

The *broad-sense heritability* is the proportion of the population trait variance attributable to variation in genotypes—a direct estimate of which is provided by the correlation between monozygotic twins reared apart (MZA). Aggregating all studies of MZA yields .75 as an estimate of the broad-sense heritability of IQ for whites reared under humane conditions (Bouchard, 1997). Studies of other kinds of twins reach a similar conclusion.

Nisbett claims that this estimate is upwardly biased, but none of his arguments withstands any serious scrutiny.

1.1.1. Selective placement

Nisbett argues that similar rearing environments induced by selective placement have inflated the correlations between the IQ scores of MZA: “[W]hen twins reared apart are brought up in highly similar environments, the correlations between their IQ scores range from .83 to .91. . . . When environments are dissimilar to one degree or another, the correlations range from .26 to .67” (Nisbett, 2009, p. 26).

Nisbett's sources on this point divided the available cases of MZA into groups deemed to differ in the similarity of their rearing conditions and calculated the IQ correlations separately in each group. These data-snooping exercises have turned out to be spurious (Bouchard, 1983). When the same grouping criterion is applied to twins from different studies or to a different IQ test used in the

same study, the differences in correlations mentioned by Nisbett simply do not replicate. Trawling through a dataset and calculating correlations for every plausible grouping scheme is unlikely to produce sound inferences.

Moreover, the same grouping criteria fail to explain the twin similarity observed in subsequent MZA studies, including the Minnesota Study of Twins Reared Apart (MISTRA). Nisbett states that “[s]ince we do not know just how dissimilar the environments are in most studies of twins reared apart, we cannot know exactly what heritability to estimate from the correlation between them” (Nisbett, 2009, p. 26), but this is not a fair summary. As a matter of fact, the MISTRA investigators collected extensive data on the rearing environments of their separated twins. While it is true that selective placement seems to have induced positive correlations between some aspects of the MISTRA twins' rearing environments, these aspects must exert a causal effect on IQ in order to act as confounding variables (Fig. 1).

Differential psychologists refer to the general factor common to all components of a typical IQ battery as “*g*”. On the basis of 74 monozygotic (MZ) and 52 dizygotic (DZ) twins reared apart, the MISTRA investigators estimated the heritability of *g* to be .77 (Johnson et al., 2007). Fig. 1 shows that the contribution of a confounding environmental variable to the IQ correlation of MZA is $r_{ge} \times r_{ee} \times r_{ge}$. To determine the sensitivity of the heritability estimate to selective placement, the investigators set r_{ge} equal to the sample correlation between *g* and the relevant environmental variable. This assumption is probably generous toward the hypothesis of confounding.

The confounding contributions of various environmental variables to the correlation between the *g* scores of the MISTRA twins were estimated as follows: family size (.00), rearing father occupation (.01), rearing mother occupation (.00), rearing father education (.00), rearing mother education (.00), total physical possessions in childhood home (.01), material possessions in childhood home (.00), cultural possessions in childhood home (.02), mechanical possessions in childhood home (.00), scientific possessions in childhood home (.01). Measurements were also taken with the following subscales of the Family Environmental Scale (FES): cohesion, expressiveness, conflict, independence, achievement orientation, intellectual-cultural orientation, active-recreational orientation, moral-religious emphasis, organization, control. The estimated confounding contribution of each of these variables was .00. There is really very little evidence to support Nisbett's insinuation that selective placement has seriously biased the heritability estimate of *g*. Later we will see that evidence for substantial confounding would have been anomalous in any case because of the small impact of sharing a household on the resemblance between other types of relatives (Table 1).

1.1.2. Biometrical models

Nisbett cites a meta-analysis estimating the broad-sense heritability of IQ to be .48 (Devlin, Daniels, & Roeder, 1997). This result

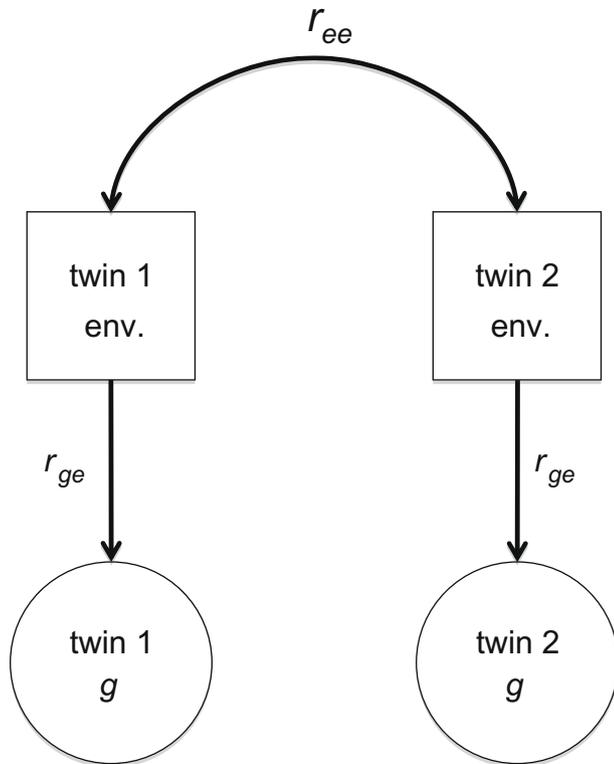


Fig. 1. Confounding effect of environmental similarity on the correlation between twins reared apart. Env. refers to some aspect of the rearing environment, g to general cognitive ability, r_{ee} to the correlation between the twins' environments, and r_{ge} to the standardized causal effect of the environmental variable on g .

Table 1
Summary of recent biometrical studies based on IQ measurements in late adolescence or adulthood.

Relationship	Correlation (95% C.I.)	Pairs	Heterogeneity p -value
<i>Adoptive relatives</i>			
Adoptive father–adopted child	.12 (.04, .20)	591	.73
Adoptive mother–adopted child	.03 (–.05, .11)	599	.17
Adoptive siblings	.05 (–.04, .14)	471	.48
<i>Parent–offspring</i>			
Biological father–reared child	.29 (.20, .37)	434	.15
Biological father–adopted child	.32 (–.03, .60)	33	–
Biological mother–reared child	.34 (.26, .42)	446	.19
Biological mother–adopted child	.36 (.26, .44)	358	.52
<i>Non-twin siblings</i>			
Half-siblings reared apart	.22 (–.02, .46)	64	–
Full siblings reared together	.42 (.31, .51)	288	.52
Full siblings reared apart	.47 (.18, .76)	28	–
<i>Twins</i>			
Dizygotic twins reared together	.37 (.28, .46)	362	.15
Dizygotic twins reared apart	.36 (.18, .51)	116	.43
Monozygotic twins reared together	.84 (.79, .88)	186	.39
Monozygotic twins reared apart	.77 (.66, .84)	87	.74

Note: These data are combined from several sources. Please email the author for references.

depends on the assumption that environmental effects on IQ endure throughout the lifespan. But a pattern of *fading* environmental effects with increasing age has been borne out in both cross-sectional and longitudinal studies (Bergen, Gardner, & Kendler, 2007; McCartney, Harris, & Bernieri, 1990; McGue, Bouchard, Iacono, & Lykken, 1993; Plomin, Fulker, Corley, & DeFries, 1997; Scarr, Weinberg, & Waldman, 1993; Polderman et al., 2006; Segal, McGuire, Havlena, Gill, & Hershberger, 2007).

Table 1 displays a summary of results from kinship studies of older individuals reported since 1978. A relatively simple biometrical model fixing the broad-sense heritability at .75 would fit these data quite adequately.

Devlin and colleagues attributed the MZ similarity in excess of their heritability estimate (.48) to shared prenatal environment, but there exists no convincing evidence for such a variance component. In fact, one study found that as dramatic a prenatal difference as whether MZ twins share a chorion during gestation is associated with trivial differences in IQ means, variances, and covariances (Jacobs et al., 2001). Such factors as fetal position, order of delivery, and blood transfusion may even act to differentiate MZ twins rather than to increase their similarity (Price, 1950, 1978). For instance, although MZ twins are eventually far more similar than DZ twins in body weight, MZ twins show greater weight differences at birth (Wilson, 1986).

1.1.3. Genetic effects mediated by physical appearance

Nisbett claims that “[t]he correlation between identical twins overestimates heritability... because the environmental experiences of identical twins who are reared apart separately in quite different environments are highly similar since they look so much alike or have other characteristics in common that tend to elicit the same sorts of behavior from other people” (Nisbett, 2009, p. 27).

It is indeed true that genes may influence phenotype through what seem to be environmental mediators. For example, if certain people find education more enjoyable because of an inherited disposition, then we can attribute positive effects of education on their abilities to the *joint* action of genetic and environmental factors. Nisbett argues that studies of MZA illegitimately apportion environmental variance of this kind to the broad-sense heritability, but the aptness of any particular apportionment is actually a subtle issue. If the goal of a heritability estimate is a rough bound on the malleability of the trait, then any difficulty in manipulating the environmental mediator may well justify placing its influence on the genetic side of the ledger. Going back to our example, we do little harm in labeling a causal path through education a genetic one if we cannot easily sever educational choices from inherited dispositions. With increasing maturity and self-determination, individuals may even seek out environments that harmonize with their innate profile of abilities and interests (Scarr & McCartney, 1983). This kind of gene–environment correlation is a plausible mechanism for the increasing heritability of IQ with age and the diminishing traces of environments experienced earlier in life (Dickens & Flynn, 2001).

So what of Nisbett’s argument invoking mediation by physical appearance? Whether such a mechanism should be counted as genetic or environmental is entirely moot if physical appearance does not even affect IQ in the first place. A meta-analysis of studies involving 3255 total participants found a correlation between physical *attractiveness* and various IQ-type measures very close to zero (Jackson, Hunter, & Hodge, 1995). Thus, any causal effects of attractiveness on IQ must make a miniscule contribution to the correlations between relatives.

Of course, one can always speculate that unmeasured dimensions of physical appearance exert the supposed mediating effect on the IQ scores of MZA. But as long as these other dimensions remain no less mysterious than “neural efficiency” and the like, it is hard to see in what sense the heritability of IQ has been explained away or brought under human control.

1.1.4. Non-additive gene action

Nisbett argues that “there may be gene interactions that specifically make identical twins more similar but that don’t contribute much to the degree of resemblance of other relatives” (Nisbett, 2009, p. 27). Geneticists use the term *epistasis* to describe such

non-additive interaction across loci, and it no doubt exists for all polygenic traits. Epistasis provides a tempting out for those who wish to argue that the findings from MZ twins have little relevance for the transmission of cognitive ability across generations.

It is indeed true that twin studies alone cannot rule out the possibility of large epistatic variance components (Keller & Coventry, 2005). Data on additional familial relationships, however, provide a test of a model where most of the genetic variance is additive rather than epistatic (Table 1). The small impact of sharing a household on familial resemblance greatly constrains the environmental degree of freedom that makes epistasis a viable candidate for explaining the resemblance between MZ twins. Furthermore, the simplest additive model predicts that first-degree relatives should be half as similar as MZ twins, and this prediction does not seem far from the truth. Nisbett ignores these data bearing out a strong additive contribution of heredity to the resemblance between parents and offspring.

A large genetic variance in the absence of *additive* genetic variance is a rather peculiar case even in theory. Consider a trait influenced by a single locus with two alleles. In order for the genetic variance to be completely non-additive, there must be equality between the means of the two homozygotes and also the frequencies of the two alleles. In the absence of such special disorderliness and symmetry, substantial additivity must be the rule.

We can extend this notion somewhat more rigorously. Even in the presence of substantial non-additive *gene action*, population-genetic theory predicts that most of the *genetic variance* in a polygenic trait should be additive in nature. Because random fluctuations in allele frequency will lead eventually to the loss of one allele, the long-term expected frequency of a mutable, weakly selected DNA variant in a population of small effective size is very near either zero or one. The rarity of one allele at many loci tends to prevent the kind of symmetrical situation leading to non-additive genetic variance. For example, even if a given pair of loci show a strong non-additive interaction, a low frequency of an allele at one locus means that an allelic substitution at the other occurs against a nearly uniform genetic background and thus exerts a predictable effect.

These theoretical considerations were borne out in a meta-analysis showing that the difference between the MZ and twice the DZ twin correlations is centered around zero for 86 assayed physical and behavioral traits (Hill, Goddard, & Visscher, 2008). The most parsimonious explanation of this pattern is that additive genetic variance typically accounts for much of the total genetic variance.

1.2. Adoption studies

Nisbett takes aim at behavioral-genetic designs other than the study of twins – in particular, the studies of biologically unrelated adoptive relatives summarized in Table 1. But his arguments here are just as unsound.

1.2.1. Restriction of range

Nisbett correctly points out that inadequate sampling of poorer families may lead to underestimates of environmental effects, but he applies this point far too broadly.

Restriction of range is a potential problem in studies of twins as well as adoptees. It is not, however, a plausible objection to MISTRA. After correction for the Flynn Effect, the standard deviation of WAIS IQ in the MISTRA twins was 14.8 – very close to that of the norming sample. In addition, the standard deviations of the FES scores were substantially greater than those obtained in their norming sample, which should have made environmental effects more readily detectable. As previously mentioned, however, these effects were estimated to be vanishingly small. Moreover, there have been some studies of entire twin cohorts from a single coun-

try born in particular years (Benyamin, Wilson, Whalley, Visscher, & Deary, 2005; Tambs, Sundet, Magnus, & Berg, 1989). Although the twins in one of these studies were only 11 years old, the results still support a heritability of approximately .70.

Nisbett cites a paper by Stoolmiller (1999) claiming that restriction of range leads to underestimates of environmental effects in studies of biologically unrelated adoptive relatives. Even given the evidence available at the time of its writing, however, Stoolmiller's argument suffers from several flaws pointed out by Loehlin and Horn (2000). A recent study found additional evidence that undermines the general argument invoking restriction of range (McGue et al., 2007). Although adopting households in Minnesota were indeed found to be less variable in many measures of environmental quality than non-adopting households, the restriction of range did not account for the IQ correlation between adolescent siblings being higher in biological than in adoptive families. In fact, corrections for range restriction increased the correlation in adoptive families by a mere .01. The regression coefficient of offspring IQ on parental SES in 242 adoptive families was statistically insignificant and in any case estimated to be four times smaller than the corresponding coefficient in biological families. Since regression coefficients are expressed on the scales of the relevant variables, restriction of range does not affect them. The authors concluded that adoption studies provide valid estimates of IQ variance attributable to rearing environment for the "broad middle class."

Nisbett's criticism of this study is quite weak. He makes a fair point that individuals living with two or more of their own children have higher rates of college graduation than the general population. But he goes on to argue that "the nonadoptive families who participated in the study were of higher SES and were probably more stable than nonadoptive families in general. As we will see later, heritabilities for such high-SES families are substantially higher than for the population at large" (Nisbett, 2009, p. 239). This is quite a departure from what the authors actually reported. The difference in rate of college education between non-adopting mothers who participated in the study (43.8%) and non-adopting mothers who did not participate (28.6%) was the sole statistically significant difference in comparisons of participating and non-participating families with respect to the following variables: education, occupational status, percent of original parents who remained married, and the number of parent-reported behavioral disorders in eligible offspring. Additionally, in a comparison with a random sample from Census 2000, the authors found small differences in percentages of fathers (47% vs. 44%) and mothers (39% vs. 44%) with college degrees. Contrary to what Nisbett implies, there is little reason to think that a different sampling scheme would have led to substantially higher estimates of environmental effects on IQ.

The argument that broader sampling might lead to much higher estimates is also contradicted by a recent study of Korean-born adoptees quite plausibly assigned at random to American adoptive families (Sacerdote, 2007). Although the sample means of maternal education and family income were above the population mean, the sample variances were either very close to the corresponding population variances or substantially greater. The correlation between adoptive siblings (1300 pairs) was .16 for highest grade completed. Stoolmiller's model of range restriction thus appears to be highly implausible; IQ is highly correlated with years of education, and his estimate of IQ variance attributable to rearing environment exceeds .16 by more than threefold.

1.2.2. Adoption studies in France

Nisbett praises a series of French studies finding IQ differences between groups of children reared in different circumstances, arguing that such "natural experiments" provide a higher grade of evidence than "[h]eritability estimates based on correlations"

(Nisbett, 2009, p. 32). The rationale for Nisbett's claim is obscure. The French studies do not differ in any fundamental way from those summarized in Table 1, and the mean difference between extreme groups is not a more theoretically relevant quantity than the correlation over the entire range of variation. For these reasons the interpretation of these French studies should aim for maximal coherence in light of the entire literature.

Capron and Duyme (1989) obtained the IQ scores of adoptees in a 2² design crossing SES (very high vs. very low) of both biological and adoptive parents. Their finding of comparable genetic and environmental effects leads Nisbett to claim that the environmental contribution to the IQ difference between extreme social classes must be as large as the genetic one. On its face the comparability of effects does seem inconsistent with the data summarized in Table 1. But we can resolve this puzzle by noting that the descriptive statistics of parental education reported by Capron and Duyme differ slightly across the nominally identical treatments. Regression coefficients computed from these differences yield an effect on adoptee IQ of roughly 2 points per year of education obtained by biological parents and 1 point per year obtained by adoptive parents (Turkheimer, 1991). Already these estimates seem more in line with those presented in Table 1. They quite plausibly come into full agreement when one considers that the adoptees in this study were tested at age 14. The positive effect of parental education observed in these French adoptees is fully compatible with other estimates of (transient) environmental effects for this age group. If 15 percent of the population variance at this age is attributable to rearing environment, then a sufficiently reliable index of environmental quality will show a correlation with IQ of $\sim .40$. The roughly 10-year difference in educational attainment between the high- and low-SES adoptive parents must then only amount to a 2-SD difference in environmental quality in order to account for the observed effect.

The study by Duyme, Dumaret, and Tomkiewicz (1999) is an even less convincing exception to the pattern in Table 1 because the inclusion criterion for the participants was a pre-adoption environment sufficiently abusive and neglectful to warrant removal from the home by judicial order. It does not seem to be in any doubt that such conditions depress IQ scores, and behavioral geneticists routinely exclude such environments from the reach of their generalizations. The contribution of such dysfunctional homes to the population variance is unclear, but it is hopefully small and diminishing.

A French study of half-sibling pairs, one reared by a working-class biological mother and another by upper-middle-class adoptive parents, found an environmental effect on IQ of 12–15 points (Schiff, Duyme, Dumaret, Stewart, Tomkiewicz, & Feingold, 1978). Since the adoptees tested in this study were between 6 and 13 years old, a large portion of this gain would probably have faded with maturation. Moreover, in a reanalysis exploiting the small variance in occupational status among the parents, Turkheimer (1991) estimated the effect on adoptee IQ of the biological father's occupational status to be three times larger than the corresponding effect of the adoptive father's occupational status. The precise conclusions of this reanalysis are doubtful because of its tenuous assumptions regarding the biological fathers and rearing environments of the non-adopted siblings. Nevertheless it reemphasizes the point that a very large *mean* difference between the rearing conditions of two groups can easily mislead us about the relative influences of *continuous* genetic and environmental variation.

1.3. The relation between heritability and malleability

Given the goal of reducing IQ differences, what is the relevance of our knowledge regarding the heritability of this trait? Surprisingly, after an entire chapter arguing that widely accepted esti-

mates of this population parameter are much too high, Nisbett writes that its precise magnitude has absolutely no relevance at all: “the degree of heritability of IQ places no constraint on the degree of modifiability that is possible” (Nisbett, 2009, p. 38, emphasis in original). Although defensible in a contrived sense, Nisbett's claim is inconsistent with the lengths to which he goes in the preceding pages.

1.3.1. Environmental engineering

It is useful here to invoke a distinction between *locally modifiable* and *modifiable in principle*. It turns out that we do not typically call something modifiable if it is only modifiable *in principle*.

For example, when we say that Alzheimer's disease is incurable, we mean, roughly, that no current medical intervention can stop the degenerative process in the brain that leads to death in about 7–10 years. ...

We would be completely baffled if someone criticized the statement that Alzheimer's disease is incurable by saying that certain effective, though presently unknown, interventions might become available some day. Of course they might! Who would deny that? Surely, the word “incurable” does not mean “something that has no cure now, and is bound to remain without cure in all eternity and in all possible worlds.” If, *per absurdum*, it did mean that, the word would be totally useless (Sesardic, 2005, pp. 164–165).

Similarly, if what Nisbett means by “no constraint” is that heritable traits are not absolutely unchangeable – that there *might* exist some undiscovered environmental intervention capable of eliminating all individual differences – then he is not saying anything particularly interesting.

A meaningful discussion of modifiability must focus on the extent to which a trait is *locally* modifiable. How much can the manipulation of known environmental variables, within feasible limits, change the distribution of IQ? Unless this class of modifications includes factors whose values do not currently vary at all, it is precisely the complement of the broad-sense heritability that constrains what such modifications can achieve. Now some of the research discussed in Nisbett's book does point to potential modifying factors with little existing variability. But many others do not seem to be of this character.

[T]he great majority of immediate policy decisions revolve around just that set of environments for which heritability estimates have the most relevance: the existing set. Most proposed policy changes involve minor redistributions of environments within the existing range, and it is precisely regarding such changes that a heritability estimate has its maximum predictive value. For instance, one message that a high heritability coefficient can convey is that minor fiddling around with environmental factors that already vary widely within the population has poor odds of paying off in phenotypic change – and thus new ideas about environments need to be tried. Surely, this is a message of enough social and practical implication to justify continued interest in heritability and its estimation (Loehlin, Lindzey, & Spuhler, 1975, p. 99).

It is true that the complement of the heritability may underestimate the trait's liability to environmental influences in the presence of gene-environment correlation and reciprocal IQ-environment feedback (Dickens & Flynn, 2001). But if we cannot readily identify or manipulate the environmental mediators of genetic effects on IQ, then the genetic variance does not in fact contain any local modifiability.

1.3.2. Genetic engineering

Once the requisite technology for the non-destructive typing and cloning of gametes is available, prospective parents will be able to choose which of their sperm and egg cells to unite in order to constitute their offspring. Given the progress to date in the mapping of genetic variants responsible for variation in other complex traits (Weedon & Frayling, 2008), we can reasonably expect to have identified hundreds of genes affecting IQ by the time that gamete-cloning technology reaches maturity. It is rather likely that the largest effects of common variants on IQ are in fact not very large – the difference between the two homozygotes probably being a point or two at best – so it may seem that a parent who is heterozygous at a handful of loci cannot make much of a difference by choosing to pass on the enhancing alleles. But even a small increase would dramatically increase the proportion of the population exceeding the ability threshold for outstanding intellectual achievement. Control over a few dozen variants may be sufficient to raise IQ and academic achievement by magnitudes rivaling even Nisbett's most optimistic appraisals of what environmental interventions can accomplish.

Since Nisbett concedes that some of his proposals may require several generations to take effect, he should not point to the time-scale of genetic engineering as a reason for excluding it from serious consideration. The heritability of IQ is thus relevant to its malleability for precisely the same reason that the heritability of a trait in livestock is relevant to breeders interested in its selective improvement: this parameter constrains the extent to which changes in the frequencies of currently segregating alleles will increase the mean of the trait.

2. Racial differences

Nisbett does not retreat from the normally taboo subject of racial differences in IQ and scholastic achievement. Indeed, he devotes nearly half of his book to this topic. He focuses in particular on the pattern of Ashkenazi Jews and East Asians performing better than non-Jewish whites and African Americans lagging behind. Nisbett clearly does not subscribe to the view that these differences are either scientifically meaningless or beyond the reach of ethical inquiry. He also rejects the argument that the selection pressures experienced by disparate human populations *must* have been so similar as to rule out any genetically based differences in cognitive abilities.

Some laypeople I know – and some scientists as well – believe that it is a priori impossible for a genetic difference in intelligence to exist between the races. But such a conviction is entirely unfounded. There are a hundred ways that a genetic difference in intelligence could have arisen – either in favor of whites or in favor of blacks. The question is an empirical one, not answerable by a priori convictions about the essential equality of groups (Nisbett, 2009, p. 94).

Thus, in principle at least, Nisbett is committed to an empirical resolution of whether heredity is a contributor to racial differences. Any in-depth review of his book must therefore follow his lead in grappling with this troubling and long-disputed question.

Nisbett argues that the evidence points squarely toward total environmental causation of racial differences. His position is at least logically compatible with the high heritability of IQ within the white population; even a trait with perfect heritability *within* one population can still vary substantially *among* populations for environmental reasons. Thus, evidence for the high heritability of IQ within one population cannot by itself demonstrate that popu-

lation differences are also genetic in origin, although such evidence does make this inference more plausible.

In his criticism of hereditarians regarding the sources of population differences, Nisbett is on firmer ground than in his early chapter on the heritability of IQ within European populations. The evidence for a substantial genetic contribution to population differences in IQ is indeed weaker than supposed by the advocates of this hypothesis. But the absence of decisive evidence favoring a genetic contribution does not entail the truth of a hypothesis attributing the entirety of the differences to environmental causes. In my view the evidence in its totality does not support either of these hypotheses clearly enough to bring closure to this contentious issue.

Nisbett does admit that the black–white difference in SES cannot completely account for the IQ difference of approximately one standard deviation between these two American subpopulations. Although Nisbett is quite satisfied with his arguments implicating environmental non-SES factors, closer examination shows them to be far from overwhelming. (I will not review Nisbett's two speculative chapters on East Asians and Ashkenazi Jews, as the research literature on these two populations is less extensive.)

2.1. Stereotype threat

Nisbett begins by invoking *stereotype threat*, an experimental manipulation leading to lower scores by blacks when the test instructions emphasize race or ability. But it is unclear whether stereotype threat makes any discernible contribution to the black–white difference observed when standardized tests are put to their typical uses.

At this point it is useful to provide more background on the psychometric technique known as *factor analysis*. Factor-analytic models treat measured variables, such as the subtests of an IQ battery, as indicators of unmeasured quantitative variables called *factors*. (Note that the term *factor* here has a narrower meaning than when used as a rough synonym for *cause* or *variable*.) There may be an infinite number of IQ items or subtests that a psychometrician might devise, but we assume that they measure only a finite number of important ability factors. If the scores on a test could be regressed on the unobserved factor scores, the resulting regression coefficients would represent the sensitivity of the test as a measure of the respective factors. The regression coefficients in this model have come to be known as *factor loadings*. Factor analysis aims to confirm what factors are measured by a set of ability tests and the loadings of the tests on each factor.

A population difference that does not arise from common factors is said to arise from *measurement bias*. In this situation a member of the minority population with the same latent ability as a member of the majority population is expected to obtain a different observed score (Fig. 2). For example, if the mean difference between two populations in vocabulary size arises from some cultural barrier impeding the minority population's acquisition of the majority language, then members of the minority population with given latent scores on *g* and the broad verbal factor will obtain lower scores on a vocabulary test than their majority peers with equal latent scores. This form of measurement bias corresponds to different intercepts in the regression of test scores on common factors. However, differences in slopes and residual variances are also forms of measurement bias, since under such conditions observed scores continue to depend on both latent abilities and group membership.

Three studies examining the factorial nature of the black–white IQ difference have found that the difference does not arise from measurement bias (Dolan, 2000; Dolan & Hamaker, 2001; Lubke, Dolan, Kelderman, & Mellenbergh, 2003). This implies that the black–white difference is indeed a difference in very general

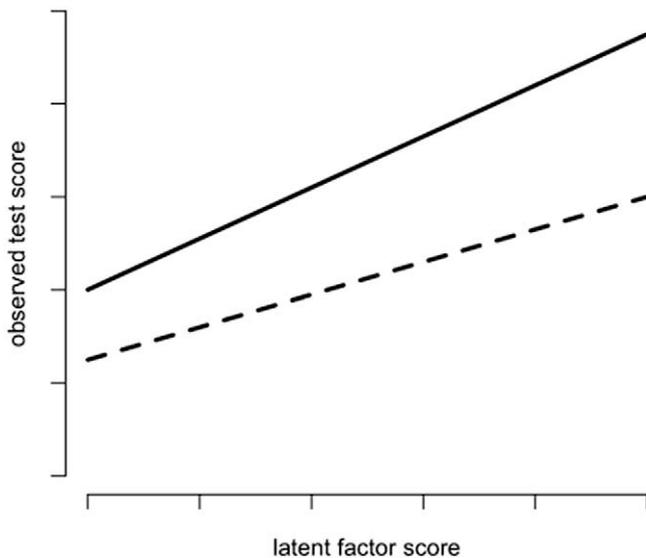


Fig. 2. Distinct factor models for two populations. The observed scores on the test exhibit measurement bias against the minority population (dashed line). For any given level of latent ability, minority members are expected to obtain lower observed scores. In this example there exists both intercept and slope (factor loading) bias.

abilities. In contrast, a study of stereotype threat employing similarly sized samples found measurement bias to be an important contributor to the differences between treatment groups (Wicherts, Dolan, & Hessen, 2005). Various experiments revealed discrepancies in intercepts, slopes (factor loadings), and residual variances, leading to the conclusion that the differences introduced by stereotype threat are generally not impairments of broad abilities. When combined with the tenability of unbiased measurement for blacks and whites in more typical settings, this finding suggests that stereotype threat may be yet another curiosity of the psychological laboratory with minimal relevance to behavior in real-world situations.

2.2. Secular trends

Raw IQ scores, much like height and other anthropometric variables, have increased in each subsequent generation over the course of the twentieth century. A reasonable extrapolation indicates that the increase has amounted to .2 standard deviations per decade. This secular increase, named the “Flynn Effect” after its discoverer, remains one of the most baffling phenomena in all of psychology. Nisbett points to the Flynn Effect as proof in principle that the black–white difference may arise entirely from environmental causes.

As Nichols (1987) pointed out in an earlier discussion of the Flynn Effect, we can characterize the argument employed by Nisbett as follows:

1. We do not know what causes the test score changes over time.
2. We do not know what causes racial differences in intelligence.
3. Since both causes are unknown, they must, therefore, be the same.
4. Since the unknown cause of changes over time cannot be shown to be genetic, it must be environmental.
5. Therefore, racial differences in intelligence are environmental in origin (p. 234).

This line of reasoning is quite treacherous. Indeed, there exists some evidence against item 3. A thorough study of the Flynn Effect’s factorial basis has shown that increases in common factors

cannot account for the entirety of the secular increase (Wicherts et al., 2005). Much as in the case of stereotype threat, the score changes constituting the Flynn Effect reflect measurement bias to some extent. “It appears therefore that the nature of the Flynn effect is qualitatively different from the nature of [black–white] differences in the United States” (p. 531).

A particular historical interaction between the Flynn Effect and the black–white difference might nevertheless strengthen the case for some relationship between the two phenomena. Specifically, have blacks in the United States experienced persistently greater generational gains than whites? Despite our ignorance of the mechanisms responsible for the Flynn Effect, an affirmative answer to this question may lead us to expect continued narrowing of the black–white gap by undirected secular trends. Nisbett claims that such a narrowing has indeed occurred and continues today.

A substantial relative increase in the black mean IQ most likely did occur during the middle of the twentieth century. The black males inducted into the armed forces during World War II may constitute the most extensive and representative sample of the black male population ever gathered, and their test data point to a black–white IQ difference at that point of 1.5 standard deviations (Loehlin et al., 1975). The specific nature of the environmental improvements responsible for the subsequent gain are unknown, although migration from the impoverished rural South to urban centers is a viable candidate.

The extrapolation of the black gain to the present day, however, is inconsistent with analyses of the National Longitudinal Survey of Youth (NLSY) and the Woodcock–Johnson IQ standardizations (Murray, 2006, 2007). The black–white difference did not diminish among NLSY children born between the mid-1970s and mid-1990s. In fact, it may have slightly increased. The Woodcock–Johnson standardizations include cohorts whose birth years span a much greater extent, and these data do show a substantial black–white convergence. This convergence ceased for cohorts born after 1970, however, and in any case narrowed an initial gap that was much greater than one standard deviation.

The overall picture that emerges from these findings is that the black–white IQ difference was approximately 1.5 standard deviations during the first half of the twentieth century, narrowed for cohorts born between 1945 and 1970 by \sim .5 standard deviations, and has remained roughly constant since that point. A similar picture emerges from the NAEP results; the black–white difference has narrowed substantially, but stagnation since the mid-1980s still leaves gaps ranging from .70 to 1.05 standard deviations (Gottfredson, 2005).

Of course, an environmentally induced shrinkage of the black–white IQ difference does not necessarily imply that the remainder is also environmental in origin. Since this remainder is factorially dissimilar to ongoing secular trends and has proven resistant to change, any pronouncements regarding its causes and future status must be treated with skepticism.

2.3. Black and biracial children reared by European parents

Studies of black and biracial children reared by European parents yield conflicting results and thus support for no particular conclusion (Table 2). Instead of acknowledging the ambiguity of these findings, Nisbett chooses to portray them as favoring his strict environmental hypothesis by launching an unprincipled attack on the study most clearly opposed to it.

The study at issue is the Minnesota Transracial Adoption Study (MTAS) (Waldman, Weinberg, & Scarr, 1994; Weinberg, Scarr, & Waldman, 1992). I first note that the MTAS data summarized in Table 2 exhibit a striking regularity that agrees with the conclusions reached earlier in this review: the environmental advantage conferred by a white upper-middle-class household for all ancestry

Table 2
Mean IQs for children of different ancestries reared by European parents.

Biological parentage	Number of children	Initial testing	10-Year follow-up
<i>Minnesota Transracial Adoption Study</i>			
Black–black	21	91.4	83.7
Black–white	55	105.4	93.2
White–white	16	111.5	101.5
Biological children	101	110.5	105.5
<i>Moore (1986)</i>			
Black–black	9	108.7	–
Black–white	14	107.2	–
<i>Eyferth (1961)</i>			
Black–white	171	96.5	–
White–white	70	97.2	–

Note: The mean IQs reflect corrections for the Flynn Effect (Loehlin, 2000). The adoptees in the Minnesota Transracial Adoption Study were initially tested at age 7. The adoptees in the study by Moore (1986) were tested between the ages of 7 and 10. The children in the study by Eyferth (1961) were tested between the ages of 5 and 13.

groups appears to be largely transient. At the second time point, both the black and white adoptees showed mean IQs extremely close to the means of their respective populations.

I also note one egregious reporting error by Nisbett. In his chapter on heritability and malleability, he cites the original MTAS publication for the following:

One study looked at the IQs of white children who were born to mothers with an average educational level and who were adopted by mostly middle- and upper-middle-class families. The children adopted relatively late had an average IQ of 117 [111.5 after correction for the Flynn Effect (Table 2)]. This study suggests that even children who would be expected to have an average IQ if raised in an average environment can have their IQ boosted very considerably if they are raised under highly propitious circumstances (Nisbett, 2009, p. 37).

Given Nisbett's extensive discussion of the later MTAS reports in his account of the black–white IQ difference, his failure to mention the longitudinal wipeout of the MTAS adoption effect is inexplicable.

At both the first and second time points, the rank order of the three adoptive groups is consistent with a hypothesis invoking some genetic contribution to the differences between the parent populations. Nisbett tries to dismiss these results by arguing that race was confounded with age at adoption, time in the adoptive home, number of prior foster placements, and quality of prior placements. This argument seems very doubtful. There exists no independent evidence that variables such as age at adoption exert effects on IQ lasting until late adolescence (van Ijzendoorn, Juffer, & Klein Poelhuis, 2005), and indeed the proportion of IQ variance associated with these pre-adoption variables declined over the course of the MTAS from .32 to .13.

The inclusion of age at adoption and the other pre-adoption variables in an analysis of covariance still managed to halve the proportion of IQ variance at age 17 associated with biological ancestry. But we should view this statistical adjustment with skepticism. Suppose that *IQ affecting pre-adoption experience* is closer to the truth than the other way around. This is quite plausible because some behavioral tendencies very early in life, including habituation to repeatedly presented stimuli and delay of gratification, can predict adult IQ with modest success (Fagan & Detterman, 1992; Mischel, Shoda, & Rodriguez, 1989). Being excitable and impulsive may very well make for a more troubled experience in foster care. Race, a plainly visible characteristic, may even more plausibly affect pre-adoption experience. In this scenario the pre-adoption variables do not act as confounders but rather as a causal fork, and their adjustment by analysis of covariance will lead to

underestimates of any true ancestry effects on IQ. For this reason the differences among the ancestry groups persisting *after* the adjustment are perhaps overly generous toward an environmental hypothesis. In summary, there is little reason to believe that the pre-adoption variables have artifactually produced what seems to be an effect of ancestry on IQ.

Nisbett concludes his case for ignoring the MTAS results with a truly remarkable argument:

Sandra Scarr [the lead investigator of the MTAS] told me that the adolescent black and interracial children had an unusual degree of psychological disturbance having to do with identity issues. Some children reported in effect, "I look in the mirror and I'm shocked to see a black person because I know I'm really white." Other children were disturbed because they felt that they were really black and did not know why they had been consigned to an alien white family (Nisbett, 2009, p. 224).

This putative mediating influence of "identity issues" is already rather suspicious because Nisbett is supposing that it exerts opposite intermediate effects in different individuals while producing the same ultimate result. Even worse is the fact that placement in a comfortable white home seems to have *raised* the IQs of the black and biracial children in the adoption study by Moore (1986). The summary of this study in Table 2 incorporates only the children reared by white adoptive parents. Both the black and biracial children reared by black adoptive parents in this study averaged substantially lower scores than their counterparts reared by whites. Unless Nisbett can explain how his hypothesis invoking identity issues can account for the opposed results obtained by Moore and the MTAS investigators, we should dismiss it as *ad hoc* speculation.

In short, studies of black and biracial children reared by white parents have so far failed to yield mutually intelligible results, and we are left with the usual proviso that more research is necessary. Nisbett's attempts to tidy up this picture carry no conviction.

2.4. "Direct" evidence of association between African ancestry and IQ

African Americans trace their origin to a relatively recent admixture of two populations that had previously evolved in isolation. Thus, African Americans can expect to inherit about 20% of their genomes from European ancestors. Nisbett points out that the hypothesis of a lower genotypic mean IQ for Sub-Saharan Africans naturally predicts that degree of European admixture should be positively associated with IQ.

Nisbett claims that the available "direct" evidence on this point supports total environmental causation of the black–white IQ difference. Despite the great weight that he attaches to them, Nisbett's sources on this point are in fact quite indecisive. He cites a study failing to find elevated European ancestry in a sample of gifted black children (Witty & Jenkins, 1936). Although this study does pose rather strong evidence for an environmental hypothesis, Nisbett does not mention a critical limitation: the investigators ascertained degree of white ancestry by parental self-report. He goes on to cite two studies failing to find an association between ancestry-informative blood-group markers and IQ without mentioning that the handicaps of small sample size and unreliable ancestry estimation rendered these two studies virtually powerless to reject any hypothesis within the interval of contention (Loehlin, Vandenberg, & Osborne, 1973; Scarr, Paktis, & Barker, 1978).

Modern genetic methodology allows estimates of ancestry admixture to draw on thousands of DNA polymorphisms rather than a mere handful of markers constrained to be associated with readily measurable phenotypic variation (Price et al., 2008). As a result we can now make such estimates with extraordinary preci-

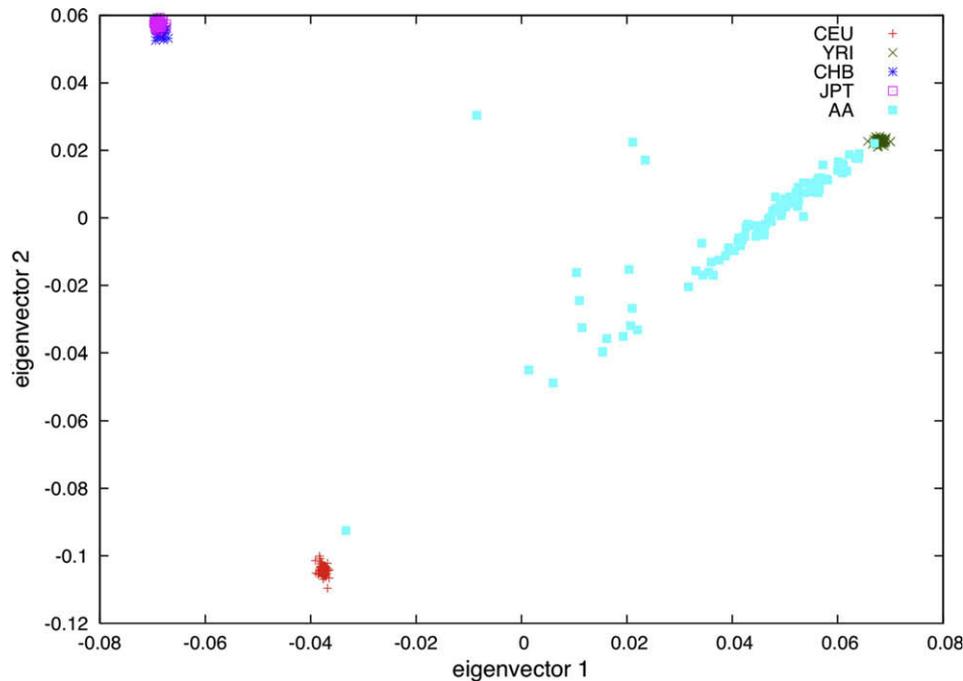


Fig. 3. Top two principal components (PCs) of ~600,000 single-nucleotide polymorphisms from 299 racially diverse Coriell and HapMap samples. European (CEU), West African (YRI), Chinese (CHB), Japanese (JPT), African American (AA). From Price et al. (2008).

sion. Fig. 3 displays what differential psychologists might call the “loadings” of several genotyped individuals on the principal components (PCs) of the genotype-by-individual matrix. We can readily see that the first two PCs perfectly separate East Asians, Europeans, and West Africans. The admixed American blacks are arrayed along a nearly straight line between the African and European clusters. The scattering toward the East Asian cluster most likely represents additional admixture with Native Americans.

If Nisbett is truly confident that degree of European ancestry shows no association whatsoever with IQ, he should call for studies employing superior ancestry estimates of the kind displayed in Fig. 3. Note that the increased reliability of ancestry estimation does not obviate the need for a large sample. Even under an extreme hereditarian hypothesis assigning mean genotypic IQs of 80 and 100 respectively to the African and European ancestors of African Americans, we can only expect an increase of .2 IQ points for every percentage increase in European ancestry. The considerable IQ variation among African Americans makes an effect of this size difficult to detect in small samples.

The ultimate test of the hereditarian hypothesis is of course the identification of the genetic variants affecting IQ and a tally of their frequencies in the two populations. Because of their likely small effects, we may have to identify dozens of such variants before we are able to make any confident inferences regarding the overall genotypic means of different populations. Although this task is currently within our technological means, it seems practically out of reach in the very short term. Ancestry estimation is much less costly than gene-trait association research and thus offers the advantage of an immediate increment toward the resolution of this issue.

3. Conclusion

Continued research with the tools of genetic epidemiology, population genetics, psychometrics, and cognitive neuroscience is likely to settle many of the contentious issues raised in Nisbett's book, even without a centralized effort toward any such narrow goal. Given that much of the critical research so clearly lies ahead, Nisbett's certainty regarding his own premature conclusions is

quite remarkable. Some of this may be owed to the disturbing possibilities raised by the alternatives. Even the prospect that current group differences might be eliminated by a combination of biological enhancement and environmental improvement will fail to put all observers at ease, since the prospect of biologically based remedies is itself frightening to many. For what it is worth, I believe that the possibilities regarding both the state of nature and our powers of control should leave us reasonably optimistic about what the future might hold. But I confess to less than total confidence in even this qualified remark, and I envy Nisbett his certitude.

References

- Benyamin, B., Wilson, V., Whalley, L. J., Visscher, P. M., & Deary, I. J. (2005). Large, consistent estimates of the heritability of cognitive ability in two entire populations of 11-year-old twins from Scottish Mental Surveys of 1932 and 1947. *Behavior Genetics, 35*, 525–534.
- Bergen, S. E., Gardner, C. O., & Kendler, K. S. (2007). Age-related changes in heritability of behavioral phenotypes over adolescence and young adulthood: A meta-analysis. *Twin Research and Human Genetics, 10*, 423–433.
- Bouchard, T. J. (1983). Do environmental similarities explain the similarity in intelligence of identical twins reared apart? *Intelligence, 7*, 175–184.
- Bouchard, T. J. (1997). IQ similarity in twins reared apart: Findings and responses to critics. In R. J. Sternberg & E. Grigorenko (Eds.), *Intelligence, heredity and environment* (pp. 126–160). Cambridge, UK: Cambridge University Press.
- Capron, C., & Duyme, M. (1989). Assessment of effects of socio-economic status on IQ in a full cross-fostering study. *Nature, 340*, 552–554.
- Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature, 388*, 468–471.
- Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review, 108*, 356–369.
- 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 analysis of the WISC-R and K-ABC and a critique of the method of correlated vectors. In F. Columbus (Ed.), *Advances in psychological research* (Vol. VI, pp. 31–60). Huntington, NY: Nova Science Publishers.

- Duyme, M., Dumaret, A.-C., & Tomkiewicz, S. (1999). How can we boost IQs of "dull children"? A late adoption study. *Proceedings of the National Academy of Sciences USA*, 96, 8790–8794.
- Eyferth, K. (1961). Leistungen verschiedener Gruppen von Besatzungskindern im Hamburg-Wechsler Intelligenztest für Kinder (HAWIK). *Archiv für die gesamte Psychologie*, 113, 222–241.
- Fagan, J. F., & Detterman, D. K. (1992). The Fagan test of infant intelligence. A technical summary. *Journal of Applied Developmental Psychology*, 13, 173–193.
- Gottfredson, L. S. (2005). What if the hereditarian hypothesis is true? *Psychology, Public Policy and Law*, 11, 311–319.
- Hill, W. G., Goddard, M. E., & Visscher, P. M. (2008). Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genetics*, 4, e1000008.
- Jackson, L. A., Hunter, J. E., & Hodge, C. N. (1995). Physical attractiveness and intellectual competence. A meta-analytic review. *Social Psychology Quarterly*, 58, 108–122.
- Jacobs, N., Gestel, S. V., Derom, C., Thiery, E., Vernon, P. A., Derom, R., et al (2001). Heritability estimates of intelligence in twins: Effect of chorion type. *Behavior Genetics*, 31, 209–217.
- Johnson, W., Bouchard, T. J., McGue, M., Segal, N. L., Tellegen, A., Keyes, M., et al (2007). Genetic and environmental influences on the Verbal-Perceptual-Image Rotation (VPR) model of the structure of mental abilities in the Minnesota Study of Twins Reared Apart. *Intelligence*, 35, 542–562.
- Keller, M. C., & Coventry, W. L. (2005). Quantifying and addressing parameter indeterminacy in the classical twin design. *Twin Research and Human Genetics*, 8, 201–213.
- Loehlin, J. C. (2000). Group differences in intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 176–193). Cambridge, UK: Cambridge University Press.
- Loehlin, J. C., & Horn, J. M. (2000). Stoolmiller on restriction of range in adoption studies: A comment. *Behavior Genetics*, 30, 245–247.
- Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence*. San Francisco, CA: W.H. Freeman.
- Loehlin, J. C., Vandenberg, S. G., & Osborne, R. T. (1973). Blood group genes and Negro-White ability differences. *Behavior Genetics*, 3, 263–270.
- 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 common factor model. *Intelligence*, 31, 543–566.
- McCartney, K., Harris, M. J., & Bernieri, F. (1990). Growing up and growing apart: A developmental meta-analysis of twin studies. *Psychological Bulletin*, 107, 226–237.
- McGue, M., Bouchard, T. J., Iacono, W. G., & Lykken, D. T. (1993). Behavioral genetics of cognitive ability: A life-span perspective. In R. Plomin & G. E. McClearn (Eds.), *Nature, nurture, and psychology* (pp. 59–76). Washington, DC: American Psychological Association.
- McGue, M., Keyes, M., Sharma, A., Elkins, I., Legrand, L., Johnson, W., et al (2007). The environments of adopted and non-adopted youth: Evidence on range restriction from the Sibling Interaction and Behavior Study (SIBS). *Behavior Genetics*, 37, 449–462.
- Mischel, W., Shoda, Y., & Rodriguez, M. I. (1989). Delay of gratification in children. *Science*, 244, 933–938.
- Moore, E. G. J. (1986). Family socialization and the IQ test performance of traditionally and transracially adopted black children. *Developmental Psychology*, 22, 317–326.
- Murray, C. (2006). Changes over time in the black-white difference on mental tests: Evidence from the children of the 1979 cohort of the National Longitudinal Survey of Youth. *Intelligence*, 34, 527–540.
- 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, 305–318.
- Nichols, R. C. (1987). Interchange: Nichols replies to Flynn. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 233–234). New York, NY: Falmer.
- Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count*. New York, NY: Norton.
- Plomin, R., Fulker, D. W., Corley, R., & DeFries, J. C. (1997). Nature, nurture, and cognitive development from 1 to 16 years: A parent-offspring adoption study. *Psychological Science*, 8, 442–447.
- Polderman, T. J. C., Gosso, M. F., Posthuma, D., van Beijsterveldt, T. C., Heutink, P., Verhulst, F. C., et al (2006). A longitudinal twin study on IQ, executive functioning, and attention problems during childhood and early adolescence. *Acta Neurologica Belgica*, 106, 191–207.
- Price, A. L., Patterson, N., Hancks, D. C., Myers, S., Reich, D., Cheung, V. G., et al (2008). Effects of cis and trans genetic ancestry on gene expression in African Americans. *PLoS Genetics*, 4, e1000294.
- Price, B. (1950). Primary biases in twin studies: A review of prenatal and natal difference-producing factors in monozygotic pairs. *American Journal of Human Genetics*, 2, 293–352.
- Price, B. (1978). Bibliography on prenatal and natal influences in twins. *Acta Geneticae Medicae et Gemellologiae*, 27, 97–113.
- Sacerdote, B. (2007). How large are the effects from changes in family environment? A study of Korean American adoptees. *Quarterly Journal of Economics*, 122, 119–157.
- Scarr, S., & McCartney, K. (1983). How people make their own environments: A theory of genotype → environment effects. *Child Development*, 54, 424–435.
- Scarr, S., Pakstis, A. J., Katz, S. H., & Barker, W. B. (1978). Absence of a relationship between degree of white ancestry and intellectual skills within a black population. *Human Genetics*, 39, 69–86.
- Scarr, S., Weinberg, R. A., & Waldman, I. D. (1993). IQ correlations in transracial adoptive families. *Intelligence*, 17, 541–555.
- Schiff, M., Duyme, M., Dumaret, A., Stewart, J., Tomkiewicz, S., & Feingold, J. (1978). Intellectual status of working-class children adopted early into upper-middle-class families. *Science*, 200, 1503–1504.
- Segal, N. L., McGuire, S. A., Havlena, J., Gill, P., & Hershberger, S. L. (2007). Intellectual similarity of virtual twin pairs: Developmental trends. *Personality and Individual Differences*, 42, 1209–1219.
- Sesardic, N. (2005). *Making sense of heritability*. Cambridge, UK: Cambridge University Press.
- Stoolmiller, M. (1999). Implications of the restricted range of family environments for estimates of heritability and nonshared environment in behavior-genetic adoption studies. *Psychological Bulletin*, 125, 392–409.
- Tambs, K., Sundet, J. M., Magnus, P., & Berg, K. (1989). No recruitment bias for questionnaire data related to IQ in classical twin studies. *Personality and Individual Differences*, 10, 269–271.
- Turkheimer, E. (1991). Individual and group differences in adoption studies of IQ. *Psychological Bulletin*, 110, 392–405.
- van IJzendoorn, M. H., Juffer, F., & Klein Poelhuis, C. (2005). Adoption and cognitive development: A meta-analytic comparison of adopted and non-adopted children's IQ and school performance. *Psychological Bulletin*, 131, 301–316.
- Waldman, I. D., Weinberg, R. A., & Scarr, S. (1994). Racial-group differences in IQ in the Minnesota Transracial Adoption Study: A reply to Levin and Lynn. *Intelligence*, 19, 29–44.
- Weedon, M. N., & Frayling, T. M. (2008). Reaching new heights: Insights into the genetics of human stature. *Trends in Genetics*, 24, 595–603.
- Weinberg, R. A., Scarr, S., & Waldman, I. D. (1992). The Minnesota Transracial Adoption Study: A follow-up of IQ test performance at adolescence. *Intelligence*, 16, 117–135.
- Wicherts, J. M., Dolan, C. V., & Hessen, D. J. (2005a). Stereotype threat and group differences in test performance. A question of measurement invariance. *Journal of Personality and Social Psychology*, 89, 696–716.
- Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., et al (2005b). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, 32, 509–537.
- Wilson, R. S. (1986). Twins: Genetic influences on growth. In R. M. Malina & C. Bouchard (Eds.), *Sports and human genetics* (pp. 1–21). Champaign, IL: Human Kinetics.
- Witty, P. A., & Jenkins, M. D. (1936). Intra-race testing and Negro intelligence. *Journal of Psychology*, 1, 179–192.

James J. Lee
 Department of Psychology,
 Harvard University,
 Cambridge, MA 02138,
 United States
 Tel.: +1 310 561 3632
 E-mail address: jameslee@wjh.harvard.edu

Available online 30 October 2009