Between-Group Mean Differences in Intelligence in the United States Are >0% Genetically Caused: Five Converging Lines of Evidence

RUSSELL T. WARNE
Utah Valley University

The past 30 years of research in intelligence has produced a wealth of knowledge about the causes and consequences of differences in intelligence between individuals, and today mainstream opinion is that individual differences in intelligence are caused by both genetic and environmental influences. Much more contentious is the discussion over the cause of mean intelligence differences between racial or ethnic groups. In contrast to the general consensus that interindividual differences are both genetic and environmental in origin, some claim that mean intelligence differences between racial groups are completely environmental in origin, whereas others postulate a mix of genetic and environmental causes. In this article I discuss 5 lines of research that provide evidence that mean differences in intelligence between racial and ethnic groups are partially genetic. These lines of evidence are findings in support of Spearman's hypothesis, consistent results from tests of measurement invariance across American racial groups, the mathematical relationship that exists for between-group and within-group sources of heritability, genomic data derived from genome-wide association studies of intelligence and polygenic scores applied to diverse samples, and admixture studies. I also discuss future potential lines of evidence regarding the causes of average group differences across racial groups. However, the data are not fully conclusive, and the exact degree to which genes influence intergroup mean differences in intelligence is not known. This discussion applies only to native English speakers born in the United States and not necessarily to any other human populations.

KEYWORDS: intelligence, IQ, group differences, behavioral genetics, race

Few constructs in the social sciences are as well understood as intelligence. One widely accepted definition states that

> Intelligence is a very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—"catching on," "making sense" of things, or "figuring out" what to do. (Gottfredson, 1997a, p. 13)

The major foundational breakthroughs in the study of intelligence occurred almost simultaneously in two different countries. Spearman's development

American Journal of Psychology
Winter 2021, Vol. 134, No. 4 pp. 480–501 • © 2021 by the Board of Trustees of the University of Illinois

of factor analysis when analyzing educational data led him to conclude that "all branches of intellectual activity have in common one fundamental function" (Spearman, 1904, p. 284), which he called "*g*" and believed was synonymous with general intelligence. The next year Binet and Simon (1905/1916) created the first successful intelligence test, which allowed users to classify examinees by their degree of intellectual disability.

In the ensuing century, psychologists and other scientists have built on these advances to create some of the most sophisticated psychometric instruments in existence and to learn more about the nature of human intelligence (Deary, 2012). As a result, psychologists know more about intelligence—its nature, causes, and consequences—than almost any other psychological construct (Warne, Astle, & Hill, 2018). For example, research in neuroscience has strongly suggested that intelligence arises from a distributed network of well-connected brain regions, located mostly in the parietal and frontal lobes (Deary, Penke, & Johnson, 2010; Haier, 2017; Jung & Haier, 2007; Penke et al., 2012). In turn, these regions—and the brain as a whole—seem to be larger, healthier, and more organized in brighter people (Flashman, Andreasen, Flaum, & Swayze, 1997; Genç et al., 2018; Gignac & Bates, 2017; W. D. Hill et al., 2019). Geneticists have identified specific genetic variances correlated with neurodevelopment and brain functioning that are also associated with interindividual differences in intelligence or educational attainment (e.g., Adams et al., 2016; Lee et al., 2018; Okbay et al., 2016). Likewise, intelligence differences have been shown to correlate with many important health, occupational, educational, and psychosocial outcomes (Gottfredson, 1997b; Jensen, 1998; Warne, 2016a), many of which intelligence tests were never designed to predict, such as longevity (Arden et al., 2016; Gottfredson & Deary, 2004).

Individual Differences in Intelligence

One consistent finding—first shown more than 90 years ago (Burks 1928/1973)—is the heritability of intelligence, which ranges between .20 and .80 in most studies conducted in industrialized countries (Plomin, DeFries, Knopik, & Neiderhiser, 2016). Heritability is the proportion of phenotype variance that is associated with genotype variance in a sample. Zero heritability indicates that a trait's variance is entirely environmental in origin, and a heritability value of 1.0 occurs when a trait's variance is entirely genetic. Almost every trait's heritability falls between these two extremes. For intelligence, age has been shown to be a moderator variable, with most studies showing that heritability increases with age (both longitudinally and in cross-sectional comparisons of cohorts), indicating that genetics have a strong influence on the development of intelligence in adulthood, whereas in childhood environment is more important (Bouchard, 2014; Mollon et al., 2021; Plomin & Deary, 2015; Plomin et al., 2016). These heritability estimates are derived from studies of twins, adoptees and their parents, and other family relationships, but these studies are not capable of identifying specific genetic variants that correlate with intelligence phenotypes. Genome-wide association studies (GWASs) are used for this purpose because they take advantage of data mining techniques to identify genetic variants that are correlated with a phenotype. Using GWASs, behavioral geneticists have thus far identified hundreds of portions of DNA that are correlated with intelligence phenotypes (Savage et al., 2018). More genetic variants will probably emerge from future studies with larger sample sizes, greater statistical power, and more sophisticated genome sequencing technology (Murray, 2020; Plomin & von Stumm, 2018).

Although the evidence of a genetic influence on interindividual differences in intelligence is strong, this is not to say that the environment is not important. No study has ever shown a heritability of 1.0 for intelligence, and no theorist expects such (Plomin et al., 2016). This leaves room for environmental influences, especially during childhood. Progress in identifying specific environmental influences on intelligence is not as advanced as the research on genetic influences. The lower level of certainty about environmental impacts on intelligence is not caused by a lack of trying to discover these influences. Rather, there are three main impediments. First, there is no strong theory about the nature of environment, how influence is transmitted from one generation to another, or how to measure it (Hunt, 2011). On the other hand, biology has clear answers (either empirically or theoretically) about these issues for genes.

A second difficulty is that the important environmental influences on intelligence seem to be mostly

nonshared influences within families. In other words, the major environmental determinants of intelligence seem to be unique individual experiences and not any influence that family members share as they live and grow up together (Plomin et al., 2016). The idiosyncrasies of these components of the environment make these causes of intelligence differences difficult to isolate, although some have tried to do so (e.g., Asbury et al., 2016).

A final difficulty in studying environment is that many so-called environmental variables are actually partially genetically caused (Kendler & Baker, 2007). This is especially true for people raised by one or both biological parents; these people share genes with their parents, and those genes, in turn, shape the home environment via parental actions (Plomin et al., 2016). This means that many of the usual explanations of environmental influence on intelligence—such as socioeconomic status or nutrition—are confounded with genetic influence. Setting aside genetic impacts on the home environment, individuals' genes help them choose an environment that is suited to their (genetically influenced) phenotypes, a tendency that accelerates as individuals have more freedom to choose environments. This may explain why genetic influences on intelligence tend to increase in adulthood (Bouchard, 2014). Despite these genetic influences on intelligence and the fact that the exact environmental variables that are important for explaining individual differences in intelligence are not fully known, it is clear that genes are not the only influence on intelligence and that environmental characteristics do matter (Gottfredson, 1997a; Neisser et al., 1996; Warne, 2020b).

Between-Group Differences in Intelligence

Just as for individual-level differences, researchers widely recognize that racial and ethnic groups differ in their mean level of intelligence (Neisser et al., 1996; Warne et al., 2018), although there is a great deal of overlap between the distributions of intelligence in these various groups (Gottfredson, 1997a; Warne, 2020a). Average score gaps in IQ and similar measures persist in the United States, with the average IQ scores being approximately 105 for Asian Americans, 100 for White Americans, 90 for Hispanic Americans, and 85 for African Americans (see Curran & Kellogg, 2016; Dickens & Flynn, 2006; Hunt, 2011; Little, 2017; Murray, 2007, 2020, 2021 for more exact estimates aggregated from representative samples).

Although knowledge of the causes of individual differences in intelligence is increasing every year, the information on causes of mean group differences in intelligence is not as advanced. This discrepancy is at least partially due to the fact that the topic of race differences in intelligence is one of the most controversial in all of science (Cofnas, 2016; Reynolds, 2000). Indeed, because of the social sensitivity of the topic, some have argued that the very act of investigating causes of these differences is itself inherently immoral or racist (e.g., Gould, 1981; Kourany, 2016; Sternberg, 2005; Turkheimer, 2007). I reject this censorious perspective as being antithetical to science. After all, "Without a forum for controversy, controversy will not be resolved and science will not advance" (Detterman, 2006, p. iv). Moreover, ceding the debate about sensitive topics leaves a vacuum for uninformed, extremist, or unscientific perspectives to take hold (Reich, 2018).

Therefore, in this section of the article I present five lines of evidence that indicate that within the United States, average differences in intelligence between racial and ethnic groups are not 100% environmental in origin. This evidence contradicts two viewpoints that are found in the literature. The first is that average differences in intelligence test scores across racial groups are purely environmentally caused (e.g., Helms, 1992; Ogbu, 2002). The second viewpoint is that there is no evidence supporting a genetic influence, so the position that average IQ differences are fully caused by environmental differences should be the default belief (e.g., Kaplan, 2015; Nisbett et al., 2012b; Sternberg, 2005). Scientists holding these viewpoints reach the same conclusion—that intergroup average differences in intelligence are environmentally caused—via two different perspectives. The first is more extreme and states that environmental differences can explain the entire mean IQ gap between racial groups; the latter is a moderate position that argues that genetic evidence has not been found (or cannot be) or that the genetic evidence simply does not exist. This article will review the evidence that both viewpoints are incorrect.

None of these lines of evidence is incontrovertible proof of a genetic influence of between-group differences in intelligence. However, the five together

provide strong circumstantial evidence that points to the same conclusion: a nonzero genetic influence on racial and ethnic groups' mean intelligence differences. Although there are other types of evidence that some advocates of a genetic cause for mean group differences find convincing, I have limited myself to the evidence that I believe is most difficult to explain with a theory of purely environmental causes of average score gaps between demographic groups.

Before I describe these five lines of evidence, limits to this discussion need to be established. First, in this article I discuss only research on American populations. The reason for this is pragmatic: There has been more research on this topic in the United States than in any other nation. It may not be justifiable to generalize the research in this discussion to other countries or on a worldwide scale. Second, when members of different groups are discussed, it is assumed that each group consists solely of native English speakers born in the United States. This is because intelligence tests are designed for people who fit this demographic profile, and professionally designed intelligence tests are unbiased against these examinees (Jensen, 1980a; Reynolds, 2000). This eliminates from this article any discussion of language barriers, large cultural differences, or unfavorable environmental influences that originate in other nations (e.g., famine, civil unrest, a low-quality school system in immigrant examinees' nations of birth).

Note About Racial Groups and Classifications

It is important to understand the nature of racial groups to avoid misunderstandings or oversimplifications. A modern understanding from population genetics defines races as ancestral breeding populations that had their recent evolutionary origin in a particular geographic region of the world (Reich, 2018). These groups are defined in terms of the differences in the relative frequencies of a large number of alleles in the genome. These genetic differences arose because of relative isolation that the groups had from one another as natural selection or genetic drift occurred over the generations. As a result of this separation in recent evolutionary history, DNA variants can be used to identify the populations that a person's recent ancestors originated from. This information aligns almost perfectly with self-identified racial and ethnic group classifications. For example, in one typical study, atheoretical DNA-based classification matched self-identified race and ethnicity with 99.86% concordance (Tang et al., 2005). Results such as these should put to rest any claims that race is purely a social construct with no grounding in biology (see also Murray, 2020). Most of the research I discuss in this article uses self-identification to classify people into groups, although some of the genomic studies use DNA-based classification. Either method produces the same results for almost all individuals.

However, it is essential to recognize that these populations in recent evolutionary history are all the product of ancient admixtures of lineages that had previously split from the basal human population (Reich, 2018). Racial groups are not static, unchanging populations and never have been. Rather, current groups are identifiable at a particular point in human evolutionary history; analyses of DNA from ancient humans produces racelike groupings that do not align with modern racial groupings. These ancient genetic populations intermixed, branched off, or merged with other populations to produce modern racial groups. As a result, modern racial groupings make sense for contemporary populations, but there is nothing enduring or unchanging about any group, and there is no genotypic or phenotypic characteristic that all individuals within a racial group share and that is missing from all non–group members (Reich, 2018). However, none of this changes the fact that modern racial groups are identifiable via genomic analysis and that *average* genetic differences are easily observable when we compare sample of living humans who are descended from recent ancestral populations who lived in different regions of the world.

Evidence 1: Spearman's Hypothesis

Based on the approximate mean scores mentioned previously, the average intelligence difference between racial and ethnic groups in the United States ranges from about 5 to 20 IQ points, or .33 to 1.33 *SD* (Gottfredson, 1997a; Hunt, 2011), depending on which groups one chooses to compare. However, the magnitude of these differences varies from subtest to subtest, even within the same intelligence test battery (Jensen, 1980a). Psychologists have long recognized this fact (e.g., Derrick, 1920), although most did not investigate the issue. When discussing the varying mean score differences between African Americans

and White Americans on subtests, Spearman (1927, pp. 379–380) off-handedly suggested that the better measures of *g* exhibited larger score gaps. This idea stayed untested until revived by Jensen (1980b), who named it Spearman's hypothesis.

Jensen (1985) created a test of Spearman's hypothesis, called the method of correlated vectors (MCV). In this procedure, the subtest's factor loading on a *g* factor—which is a measure of the degree to which the subtest measures *g*—and the effect size measuring the Black–White mean differences are correlated with one another. MCV results when comparing racial groups often show that these two vectors are positively correlated, indicating that Spearman was correct that better measures of *g* tend to have wider mean score differences between groups. These correlations vary but are consistently positive (e.g., $r = .84$ in Dahlke & Sackett, 2017; $r = .59$ in Jensen, 1985; $r = .46$ in Kane, 2007; $r = .57$ in te Nijenhuis & van den Hoek, 2016; $r = .723$ to $r = .817$ in Warne, 2016b). When expanded to other comparisons, the MCV seems to support Spearman's hypothesis when researchers compare White and Hispanic groups (Dahlke & Sackett, 2017; Hartmann, Kruuse, & Nyborg, 2007; Kane, 2007; te Nijenhuis, van den Hoek, & Dragt, 2019; Warne, 2016b) and White and Native American groups in the United States (te Nijenhuis, van den Hoek, & Armstrong, 2015). Results for Asian–White MCV analyses are inconsistent (Kane, 2007; Nagoshi, Johnson, DeFries, Wilson, & Vandenberg, 1984; Warne, 2016b).

The importance of Spearman's hypothesis for the study of causes of mean racial group differences in intelligence test scores is not obvious at first. However, if Spearman's hypothesis is true, then it would indicate that within- and between-group differences in intelligence arise from many of the same sources (genes and environment) because the sources of within- and between-group variation are positively correlated and share variance with one another. This finding would be consistent with the belief that group mean differences in intelligence are mostly (or perhaps entirely) the product of the accumulation of individual differences (Jensen, 1998; Rushton & Jensen, 2005). Moreover, it would indicate that any unique environmental influences that operate on one group would be trivial in their impact on between-group differences.

A strength of the Spearman's hypothesis literature is that it is a subset of a larger body of literature using the MCV. These studies consistently show that subtest *g* loadings correlate positively with other biological characteristics of the brain and nervous system (Jensen, 1998, Chapter 6). More tellingly, though, is the finding that the MCV produces positive correlations with phenomena that are purely genetic, such as inbreeding depression (Rushton, 1999), and zero correlations with phenomena that are purely environmental in origin, such as the Flynn effect (Rushton & Jensen, 2010), adoption (te Nijenhuis, Jongeneel-Grimen, & Armstrong, 2015), brain damage caused by trauma or prenatal exposure to toxins (Flynn, te Nijenhuis, & Metzen, 2014), and increased guessing on test items (Woodley, te Nijenhuis, Must, & Must, 2014). This finding provides evidence that positive correlations between *g* loadings and the magnitude of the gap in test scores between most groups in the United States probably has a partially genetic cause. MCV tests that produce zero correlations for purely environmentally caused group differences also contradict the claim that Spearman's hypothesis is a methodological artifact (e.g., Guttman, 1992).

However, it is important to recognize the limits of the MCV evidence regarding Spearman's hypothesis. One problem is that the MCV lacks specificity and that a positive correlation does not conclusively prove that mean differences between groups are due to differences in *g* or the same causes as differences in *g*. This shortcoming arises because the method does not detect violations of statistical assumptions that must be met in order to justify such conclusions (Dolan & Hamaker, 2001; Lubke, Dolan, & Kelderman, 2001; Wicherts, 2017). Additionally, MCV results are sensitive to subjective decisions about data preparation, such as the selection of subtests used to identify a *g* factor (Ashton & Lee, 2005), although collecting scores from a wide variety of subtest formats generally solves this problem (Jensen, 1992, 1998). Some statistical corrections are available to adjust for the restriction of range that can develop from a narrow selection of subtests (te Nijenhuis et al., 2019), although these rely on assumptions that may be difficult to prove. Thus, although the results of a positive correlation in an MCV analysis are consistent with Spearman's hypothesis and the belief that between- and within-group differences have the

same causes, it is not conclusive evidence. Nonetheless, the logic of the MCV is attractive, and when data meet the necessary statistical assumptions and results are based on meta-analyses instead of single studies, Spearman's hypothesis provides evidence in support of the claim that mean group differences in intelligence are not fully environmental in origin.

Evidence 2: Tests of Measurement Invariance

In response to the inherent weaknesses of the MCV, some people in the intelligence research community began applying tests of measurement invariance to the study of group differences in intelligence. Measurement invariance was created by Meredith (1993) in order to strengthen the comparison of psychometric scores across demographic groups. The procedure uses multigroup confirmatory factor analysis (CFA) to test whether the necessary and sufficient conditions for comparing scores (the same factor structure, mean structure, and factor loadings for all groups) are present in the data derived from different groups. When these conditions are met, then mean differences in scores across demographic groups of examinees are due to actual differences in the construct being measured (Meredith, 1993). Since the invention of tests of measurement invariance, psychologists have applied the technique to intelligence test data many times. These results usually show that the factor structure and other psychometric properties of intelligence test data show a great deal of consistency across demographic groups (e.g., Beaujean, McGlaughlin, & Margulies, 2009; Li, Sano, & Merwin, 1996; Maller, 2000; Wicherts et al., 2004).

Dolan (2000) and his colleagues (Dolan & Hamaker, 2001; Lubke et al., 2001) realized that measurement invariance could overcome the problems inherent in the MCV. For example, measurement invariance procedures explicitly test the assumptions needed to justify the conclusion that between-group differences in *g* are due to the same causes as within-group differences in *g*, a characteristic that the MCV lacks (Dolan & Hamaker, 2001; Lubke et al., 2001). Moreover, tests of measurement invariance are an application of multigroup CFA (Meredith, 1993), unlike the MCV with its foundation in exploratory factor analysis. This means that the advantages of CFA, such as measures of model fit and the ability to test competing models, are intrinsic properties of tests of measurement invariance. This permits more complex and sophisticated investigations than are possible with the MCV and eliminates the shortcomings that can accompany practical applications of the MCV. For example, Frisby and Beaujean (2015) could test Spearman's hypothesis with a nonhierarchical factor structure—something that is not possible with the MCV.

More importantly, though, if measurement invariance holds across groups, then individual and mean group differences in a variable *must* be due to the same causes (Dalliard, 2014; Lubke, Dolan, Kelderman, & Mellenbergh, 2003). Therefore, tests of measurement invariance can eliminate the possibility that an influence on a phenotype operates on one group and not another. This is a stronger conclusion than is permitted by the MCV procedure. Research on intelligence test scores in the United States shows that measurement invariance holds across racial groups. Indeed, conducting tests of measurement invariance is typical practice in the development of professionally designed tests because it indicates the absence of test bias (Warne, Yoon, & Price, 2014), which test creators must eliminate to meet professional ethical standards for test development (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014).

To sum up, when measurement invariance holds, intelligence tests measure the same construct in the same way for different demographic groups. Measurement invariance also indicates that the same environmental and genetic causes of individual differences in intelligence within racial groups also determine between-group mean differences. However, it is important to note that this does not necessarily indicate that heritability values (whether within different groups or between groups) will be equal. Measurement invariance also does not conclusively prove that between-group heritability is greater than zero because mean environmental differences may be so large that between-group heritability could be zero. The discussion regarding Evidence 3 expounds this idea.

Evidence 3: Mathematical Relationship of Within- and Between-Group Differences

The degree of heritability of intelligence within various human groups and between groups is an empirical question. As stated previously, most heritability

estimates fall between .20 and .80. Most of these estimates are derived from samples of White individuals, although in a recent meta-analysis, the average heritability of intelligence for White, African American, and Hispanic American samples was between .58 and .73 (Pesta, Kirkegaard, te Nijenhuis, Lasker, & Fuerst, 2020, Table 3), indicating that heritability of IQ has a degree of stability across the largest racial groups within the United States.

Despite the heritability of intelligence within groups, many writers claim that within-group heritability provides no information about between-group heritability (e.g., Neisser et al., 1996), often using a thought experiment popularized by Lewontin (1970). In this analogy, one could take two random handfuls of seeds, planting one handful in fertile, moist soil. The other handful would be planted in barren, dry soil. After the seeds sprout and grow, any resulting phenotypic differences within either group must be completely genetic because the seeds within each group all experienced the same environment (i.e., all remaining differences are due solely to genetics). Conversely, all mean differences between groups must be solely environmental because the assignment of seeds to groups was random, and any genetic differences between groups have been balanced out.

Lewontin's analogy is flawed when applied to humans because humans are not randomly assigned to demographic groups (Warne et al., 2018). As a result, group mean differences in a phenotype probably are at least a partial consequence of aggregated individual differences in the phenotype. In turn, the individual differences are often partially caused by individual genetic differences.

DeFries (1972, p. 9) derived the mathematical relationship of between-group heritability and within-group heritability:

$$h_b^2 = (h_w^2)\frac{r_g(1-r_p)}{r_p(1-r_g)},$$

where

> h_b^2 is the heritability between groups (i.e., the proportion of between-group phenotype variance that is due to genetic differences between groups),
>
> h_w^2 is the heritability within groups,
>
> r_g is the genetic intraclass correlation within groups (which is a measure of genetic relatedness of group members), and
>
> r_p is the phenotypic intraclass correlation within groups (i.e., a measure of phenotypic similarity within a group).

The existence of the equation alone shows that the claim that "the genetic basis of the difference between two populations bears no logical or empirical relation to the heritability within populations" (Lewontin, 1970, p. 7) is incorrect. In the scenario of two randomly formed groups of organisms (or genetically identical organisms raised in separate environments), $r_g = 0$, which makes the entire right side of the equation equal to zero, indicating that $h_b^2 = 0$ for randomly formed groups (DeFries, 1972). This validates Lewontin's analogy; where Lewontin erred is in applying the special case of randomly formed groups to all human demographic groups, which by definition have r_g values greater than zero.[1]

Jensen (1998, pp. 447–458) applied DeFries's (1972) equation and showed that when $r_g > 0$, h_b^2 increases as h_w^2 increases, indicating that unless members of a racial group are completely unrelated to one another (i.e., $r_g = 0$, which by definition is impossible because racial group members belong to the same ancestral population), a value of h_b^2 greater than 0 is likely. However, an estimate of h_b^2 is not possible to derive because r_g is also unknown. Additionally, even with non–randomly formed groups, it is still possible for h_b^2 to be zero if between-group environmental differences are large enough to overpower the influence of between-group genetic differences (Warne, 2020b, Chapter 28).

The question then becomes whether the mean environmental differences between Black and White individuals in the United States are so large that they result in a between-group heritability (h_b^2) value of zero (Flynn, 1980). Expressed in standard deviation units, the required mean differences in environment associated with a between-group heritability (h_b^2) are expressed as Jensen's (1998, p. 454) equation:

$$\sigma_E = \sqrt{\frac{1-h_b^2}{1-h_w^2}},$$

where σ_E is the mean environmental difference between groups expressed in standard deviation units (i.e., Cohen's *d* effect size).

Whereas h_w^2 has empirically derived values, h_b^2 and σ_E remain unknown. Nevertheless, the equation can provide circumstantial evidence regarding the value of h_b^2. Armed with this equation, one can pos-

tulate various values for h_b^2 and h_w^2 and then examine whether the resulting σ_E values are plausible for a given mean between-group difference in a phenotype. Table 1 shows these calculations for a $d = 1.0$ difference in intelligence. Some of the σ_E values are inherently implausible. For example, if within-group heritability values of intelligence for African American and White American individuals are typically between 0.20 and 0.80, the values of σ_E in the top two rows (where $h_w^2 = 0.00$ and 0.10) and the bottom row (where $h_w^2 = 0.90$) are not realistic.

Table 1 illustrates an important feature of the relationship between between-group heritability and within-group heritability: When the two heritability values are equal, a 1-*SD* mean environmental difference produces a 1-*SD* difference in intelligence test scores. This is illustrated in the diagonal of 1.000 values, the cells that have equal h_b^2 and h_w^2 values. However, when $h_b^2 > h_w^2$, mean differences in environment must be less than the observed phenotype difference, as indicated by the σ_E values less than 1.000 above the diagonal. Conversely, when $h_b^2 < h_w^2$, then the mean environment differences must be greater than the mean phenotype difference, as indicated by the σ_E values greater than 1.000 below the diagonal. An implication from this table is that anyone who postulates that between-group heritability is less than within-group heritability must expect a larger mean environmental difference between groups than the mean difference observed in the actual phenotype. Additionally, the lower the postulated h_b^2 value, the larger environmental differences must be to produce the observed mean phenotype difference. Applied to intergroup intelligence differences, this implies that for between-group differences to be entirely genetic (i.e., for h_b^2 to be zero), the mean between-group environmental differences must be larger than the $d = 1.00$ mean difference observed between African American and White American individuals. Empirical data about the size of these differences can help judge the plausibility of different h_b^2 values, including whether h_b^2 could be equal to zero.

"Environment" is a catch-all term for any nongenetic effect on a phenotype (Hunt, 2011). Still, relevant environmental influences on intelligence differences (whether within or between groups) must be causal in nature. Therefore, any discussion of environmental influences can ignore irrelevant or noncausal environmental factors. Although this may seem like a banal insight, it does whittle down all environmental variables to a theoretically relevant set of variables that could affect a phenotype (Jensen, 1998). One of the most frequently suggested environmental causes of mean differences in intelligence among African Americans and White Americans is socioeconomic status (e.g., Nisbett et al., 2012a). Although socioeconomic status is by no means the only plausible environmental cause for mean differences in intelligence across

TABLE 1. Projected Values of Mean Environmental Differences (σ_E), Given Between-Group Heritability (h_b^2) and Within-Group Heritability (h_w^2) Values for a Phenotype (IQ) Difference of $d = 1.00$

h_w^2	h_b^2 values										
values	.00	.10	.20	.30	.40	.50	.60	.70	.80	.90	1.00
.00	1.000	0.949	0.894	0.837	0.775	0.707	0.632	0.548	0.447	0.316	0.000
.10	1.054	1.000	0.943	0.882	0.816	0.745	0.667	0.577	0.471	0.333	0.000
.20	1.118	1.061	1.000	0.935	0.866	0.791	0.707	0.612	0.500	0.354	0.000
.30	1.195	1.134	1.069	1.000	0.926	0.845	0.756	0.655	0.535	0.378	0.000
.40	1.291	1.225	1.155	1.080	1.000	0.913	0.816	0.707	0.577	0.408	0.000
.50	1.414	1.342	1.265	1.183	1.095	1.000	0.894	0.775	0.632	0.447	0.000
.60	1.581	1.500	1.414	1.323	1.225	1.118	1.000	0.866	0.707	0.500	0.000
.70	1.826	1.732	1.633	1.528	1.414	1.291	1.155	1.000	0.816	0.577	0.000
.80	2.236	2.121	2.000	1.871	1.732	1.581	1.414	1.225	1.000	0.707	0.000
.90	3.162	3.000	2.828	2.646	2.449	2.236	2.000	1.732	1.414	1.000	0.000

American racial groups, it does produce a useful test case for the claim that between-group heritability of intelligence is zero. The U.S. Department of Education's National Center for Educational Statistics has made data publicly available from several representative longitudinal samples of children. The mean differences in socioeconomic status—a composite variable created by department personnel based on parental income, occupation, educational level, and (in one sample, the High School Longitudinal Study) the urbanicity of the child's high school—in the initial year of each study are shown in Table 2. The mean differences between African American and White American children's family socioeconomic status ranges from $d = 0.604$ to $d = 0.771$, with a weighted average of $d = 0.658$. All these values fall far short of the $d = 1.118$ to $d = 2.236$ environmental differences needed to produce a between-group heritability value of zero. Given within-group heritability of .20 to .80, these socioeconomic status differences account for only 7.3% to 47.6% of the environmental variance needed to produce an h_b^2 value of zero.[2]

No theorist advocating for an h_b^2 value of zero would argue that the environmental impacts on intelligence are captured entirely by childhood socioeconomic status. The fact that the Cohen's d values in Table 2 are not large enough to result in an h_b^2 value of zero is not damning in any way to the claim that between-group differences in intelligence are fully environmental. It is likely that other environmental variables contribute to between-group mean differences in intelligence between African Americans and White Americans. However, the difficulty then becomes generating a list of variables that have a known causal impact on intelligence test scores and are nonredundant with socioeconomic status and any other environmental variables that cause between-group differences (Flynn, 1980). With the unimpressive results of experimental efforts to permanently raise IQ scores among people in industrialized countries (e.g., Lipsey, Farran, & Durkin, 2018; U.S. Department of Health & Human Services, 2012) and the large reduction in effect sizes in correlational studies after for genetic confounds were controlled for (e.g., Bouchard, Lykken, Tellegen, & McGue, 1996), finding environmental variables that fit these requirements and that can account for the remaining 52.4% to 92.7% of environmental variance needed to produce an h_b^2 value of zero seems implausible (Warne et al., 2018). If the correlation of socioeconomic status with intelligence is not due to a causal impact of socioeconomic status on intelligence, then the amount of variance that must be explained by other environmental variables for h_b^2 to be zero increases. In other words, it may not be enough for other environmental variables to explain 52.4% to 92.7% of variance in intelligence for h_b^2 to be zero.

However, the average socioeconomic differences between African American and White American children can also help eliminate some of the heritability values in Table 1 from consideration. If one accepts the weighted average of $d = .658$ as the best estimate of observed mean socioeconomic status differences between African Americans and White Americans, then any combination of h_b^2 and h_w^2 values that produces a needed environmental difference less than $d = .658$ is inherently implausible because such environmental differences would produce mean IQ score dif-

TABLE 2. Mean Differences in African American and White American Socioeconomic Status in Representative Samples of Schoolchildren

Dataset	Child grade at data collection	Mean difference (Cohen's *d*)
National Educational Longitudinal Study of 1988	8	0.625
Early Childhood Longitudinal Study, Kindergarten Class of 1998–1999	Kindergarten	0.771
Education Longitudinal Study of 2002	10	0.606
High School Longitudinal Study of 2009	9	0.604
All samples		0.658 (weighted average)

ferences larger than $d = 1.0$. in Table 1. This eliminates scenarios in which high values of between-group heritability are coupled with low values of within-group heritability, such as $h_b^2 = 0.70$ and $h_w^2 = 0.20$, which would need only an environmental difference of $d = 0.612$ to appear.

The information in Tables 1 and 2 and the mathematical constraints on the relationship that exists for between-group heritability and within-group heritability are powerful circumstantial evidence that the causes of mean differences in intelligence between African Americans and White Americans are not completely environmental. However, the information in this section is not exact enough to produce a likely value of between-group heritability.

There are two further difficulties for a purely environmental explanation for average IQ differences between races: measurement error and genetic influences on environment. Measurement error attenuates linear correlations and therefore requires that observed group differences in environment variables be larger than reported in Table 1 in order to reduce h_b^2 to zero. For example, if one believes that $h_w^2 = .50$, then the needed mean environmental difference between groups is $d = 1.414$ if environmental variables and intelligence are measured without error. However, if reliability for both variables is .90, then the needed observed difference becomes

$$d = \frac{1.414}{\sqrt{r_{xx'} r_{yy'}}} = \frac{1.414}{\sqrt{(0.90)(0.90)}} = \frac{1.414}{\sqrt{0.81}} = \frac{1.414}{0.90}$$

$$= 1.571.$$

Lower reliability values will drive the needed observed effect size even higher. As a comparison, Herrnstein and Murray (1996, p. 598) reported a Cronbach's α value of .76 for their measure of socioeconomic status and at least .90 for their measure of intelligence. Using the equation above implies that with these reliability values, an observed mean environmental difference of

$$d = \frac{1.414}{\sqrt{(0.76)(0.90)}} = \frac{1.414}{\sqrt{0.684}} = \frac{1.414}{0.827} = 1.710$$

between groups is needed to produce an h_b^2 value of zero when is h_w^2 .50 for a latent, error free measure of intelligence.

The second difficulty of genetic influences on environment arises from the fact that most people live in environments that are at least partially influenced by their genes, which means that even an "environmental" variable is often partially caused by genetics (Kendler & Baker, 2007; Plomin et al., 2016). Controlling for genetic confounds in environmental variables will correct the overestimated explanatory power of the environmental variables in correlational data and make it even more difficult for these variables (e.g., socioeconomic status) to explain between-group mean differences in intelligence.

Evidence 4: Data From GWASs

The first three forms of evidence are all what James Flynn (1980) classified as indirect forms of evidence regarding the possible influence of genes on intergroup average differences in intelligence. All three rely on statistical and methodological characteristics of data to show that purely environmental explanations of mean group differences are implausible. Though useful, these types of evidence do not provide evidence about the functioning of genes themselves. Instead, they focus on the observed patterns of data and use them to infer the probability of a genetic impact. In contrast, direct evidence examines the impact of genetics and is stronger than indirect evidence because it relies on fewer untested assumptions about the environment or genes and produces strong inferences about how group mean IQ differences would change if environments were equalized (Flynn, 1980).

A new form of direct evidence that between-group differences in intelligence are not fully environmental comes from GWASs. Currently, hundreds of genetic variants have been identified that correlate with intelligence test scores (Savage et al., 2018). The identification of DNA segments associated with intelligence would be interesting in its own right, but once these portions of DNA are known, it is possible to create an estimate of genetic influence (based on known relevant variants) on a particular person's phenotype. This estimate, called a polygenic score, is a variable created by assigning a weight to each variant of each relevant portion of DNA. Then, the weights for the variants in a particular individual's genome are summed to create an overall polygenic score for that individual (see Plomin & von Stumm, 2018, for a nontechnical explanation with a concep-

tual example). These polygenic scores correlate with intelligence test scores modestly ($r = .228$, $r^2 = .052$), but as more DNA variants associated with intelligence are identified these correlations will increase (Plomin & von Stumm, 2018; Savage et al., 2018). It is important to note that the phenotype variance explained by specific genes identified via GWAS are lower than the total within-group heritability identified through kinship studies (5.2% vs. 20–80%). This difference is called missing heritability and is probably due to current GWASs having sample sizes too small or lack the genetic diversity needed to detect rare genetic variants that can affect intelligence and to the limitations of current technology (Plomin & Deary, 2015; Plomin & von Stumm, 2018; Yang et al., 2015). Missing heritability occurs with many other phenotypes and is not unique to intelligence.

GWASs and polygenic scores provide the key to determining whether specific segments of DNA are associated with average intelligence score differences across groups because if the same DNA segments are identified in different racial groups as correlating with intelligence or if polygenic scores are correlated with intelligence test scores for different racial groups, then the most parsimonious conclusion is that some genes determine intelligence phenotypes for multiple racial groups. This is exactly the sort of direct evidence that Flynn (1980) stated would be more convincing than the indirect evidence based on statistical and methodological inferences I discussed earlier in this article.

Thus far, the majority of GWASs have been conducted on samples that consist mostly or entirely of individuals of European descent (Popejoy & Fullerton, 2016). This limitation creates a problem because genetic variants are not uniformly distributed across human populations, because of geographic or social separations during human evolutionary history (Reich, 2018) and genetic drift. As a result, genetic variants associated with intelligence in Europeans may not generalize to other racial groups (Domingue, Belsky, Conley, Harris, & Boardman, 2015). Additionally, because they are based on the statistical principles of regression, polygenic scores face an inevitable shrinkage in explained phenotype variance when applied to an independent sample.

Despite these challenges, there are indications that polygenic scores and GWAS results from Europeans can be partially generalized to other racial groups. In one study (Domingue et al., 2015), a polygenic score calculation method for adult educational attainment calculated with genetic Europeans' GWAS data were used to generate polygenic scores for African Americans, and these polygenic scores correlated with adult educational attainment $r = .11$ ($r^2 = .012$, $p = .004$; compare the genetic Europeans' correlation of $r = .18$, $r^2 = .032$, $p < .001$). Although adult educational attainment is not the same as intelligence, the two traits are positively correlated both at the phenotypic level (Damian, Su, Shanahan, Trautwein, & Roberts, 2015; Domingue et al., 2015) and at the genetic level via shared genes (Krapohl et al., 2014; Lee et al., 2018; Okbay et al., 2016). In fact, polygenic scores for educational attainment in adults predict phenotype intelligence test scores better than polygenic scores for intelligence do (Plomin & von Stumm, 2018). Thus, the fact that polygenic scores derived from individuals of European ancestry can generalize to African Americans—albeit with some shrinkage—is direct evidence that at least some of these genes influence group differences in mean intelligence test scores.

Further evidence comes from another study of applying GWAS data derived from European samples to other populations. Piffer (2015) used known frequencies of genetic variants that in published GWASs were associated in individuals descended from Europeans with better performance on intelligence tests or other measures of general cognitive functioning (e.g., executive functioning, educational attainment). He found that the frequency of these alleles in a nation's population was strongly correlated with estimated national mean IQ ($r = .91$). The positive correlation was maintained after the degree of genetic relatedness between populations was controlled for, a finding that Piffer (2019) later replicated. Moreover, a random set of alleles had a much weaker relationship with national-level mean IQ. Just as in the Domingue et al. (2015) study, this provides evidence that some of the alleles associated with higher intelligence in Europeans also have this relationship in other groups. It is likely that at least some of these genes are responsible for mean intelligence differences across racial groups.

It is important to recognize the limitations of this line of evidence. One is that, compared with other continents of origin, people with ancestors from Eu-

rope are not a very genetically diverse population (The 1000 Genomes Project Consortium, 2015), which may limit the ability of a GWAS to detect genetic variants associated with a phenotype if those alleles are comparatively rare in Europeans. Another problem is that polygenic scores and other GWAS results derived from Europeans are less generalizable to other populations as the degree of genetic relationships with Europeans decreases. As a result, applying results of GWASs to other populations sometimes produces nonsensical or incorrect results (Martin et al., 2017).

Nevertheless, the results of the Domingue et al. (2015) and Piffer (2015, 2019) studies are evidence that mean intelligence score differences across racial groups are not entirely environmental in origin. In the future, GWAS samples are likely to become more diverse because the health benefits of identifying alleles associated with phenotypes (e.g., for heart disease, diabetes) should be extended to all human populations (Popejoy & Fullerton, 2016; Reich, 2018). If this research progresses as it did for Europeans (which is likely), researchers conducting some GWASs will collect intelligence scores—or scores for highly correlated variables, such as educational attainment—from these diverse samples. This research will probably have two consequences. First, researchers will identify new genetic variants associated with intelligence, which will improve polygenic scores. Second, polygenic scores derived from more diverse samples will be more generalizable (Martin et al., 2017). These developments, already under way, will also provide stronger evidence of whether intergroup mean differences in intelligence are partially genetically caused.

Evidence 5: Admixture Studies

Another form of direct evidence that sheds light on whether between-group differences are fully environmental in origin is admixture studies. This method studies admixed populations, which are groups that are descended from ancestral populations that are native to different parts of the world (often different continents). In admixture studies, researchers capitalize on the individual variability in ancestry source within the admixed population by correlating the proportion of individuals' ancestry that originates from a particular continent with a phenotype score. If this correlation is not zero, then this is evidence that the trait has a degree of between-group genetic influence (Dalliard, 2014; Martin et al., 2017). Modern genomic-based admixture studies originated in medical research and have been successful at identifying health phenotypes that are associated with genetic heritage from different continents (e.g., Cheng et al., 2012; Flores et al., 2012; Kao et al., 2008).

In the United States, the largest admixed populations are Hispanic Americans and African Americans. Among Hispanic Americans, an average of 55–70% of ancestry originated in Europe (Bryc, Durand, Macpherson, Reich, & Mountain, 2015; Kirkegaard, Woodley of Menie, Williams, Fuerst, & Meisenberg, 2019), with the rest originating in the Americas and Africa. For African Americans, an average of 15–25% of ancestry originated in Europe, with the remaining ancestry originating almost entirely in Africa (Bryc et al., 2015; Jin et al., 2012). Within these groups of admixed individuals, there is a positive correlation (r = .23 to .30) between intelligence and the degree of European ancestry an individual possesses (Hu, Lasker, Kirkegaard, & Fuerst, 2019; Kirkegaard et al., 2019; Lasker, Pesta, Fuerst, & Kirkegaard, 2019; Warne, 2020a). This correlation cannot be explained by within-group discrimination based on skin color (Hu et al., 2019; Krieger, Sidney, & Coakley, 1998; Lasker et al., 2019), and controlling for socioeconomic status does not reduce the correlation to zero (Lasker et al., 2019). A purely environmental explanation of group differences in average intelligence has difficulty explaining the correlations between European ancestry because the correlations, by definition, have a genetic component.

Admixture studies of intelligence based on genetic assertations of ancestry are a recent phenomenon, and thus far only four have been published. However, with ancestral DNA tests becoming widespread and the growing push to incorporate genomic data into the social sciences (e.g., Murray, 2020; Plomin, 2018), more admixture studies will probably follow. In addition to replications of these early studies, future admixture studies on other groups will be an important theoretical test of the hypothesis of mean differences in intelligence across groups having a genetic influence because admixture studies lead to testable hypotheses about the possibility of a genetic influence on between-group average differences in intelligence. For example, Warne (2020a) extrapolated from admixture data in African Americans

and Hispanic Americans to predict that individuals with European heritage admixed with ancestry from groups with higher average scores on intelligence tests, such as East Asians, should have a *negative* correlation between the proportion of their European ancestry and their intelligence test score. It is plausible that Warne's (2020a) hypothesis will be supported in such admixture tests; in one study of White Americans, individuals with ancestry from European Jews, the ethnic group with the highest average intelligence test score in the world, had higher polygenic scores for intelligence than sample members from other European groups (Dunkel, Woodley of Menie, Pallesen, & Kirkegaard, 2019). If admixture studies of European Jews and East Asians show a negative correlation between intelligence test scores and non-Jewish European ancestry, then it would indicate that people with high intelligence are more likely to have, on average, more genetic variants associated with higher intelligence, plus some environmental advantages (Warne, 2020a).

Still, correlation—even a genetic correlation—is not causation. One well-known objection to the statement that genetic correlations must have a genetic cause is from Jencks (1979), who stated that if a heritable trait is correlated with a proximal environmental cause of group differences, then it can result in a specious correlation between genes and an environmental outcome. Jencks proposed a thought experiment where children with red hair are systematically denied access to schooling; as a result, their academic performance would be correlated with the alleles for red hair, but that would not indicate that the alleles that encode for red hair have a causal impact on academic performance. However, no one has ever shown that such arbitrary environmental deprivations actually occur in 21st-century America. A hypothetical scenario where a particular gene–phenotype correlation was spurious does not mean that all—or any—such correlations in the real world are spurious. The parsimonious explanation for a correlation between genetic variations and a psychological phenotype is that genetic variance is a partial cause of phenotype variance. Scientists who postulate an interpretation that a particular correlation is an artifact of differential environments have the burden of proof of producing data that such environmental differences across genetic groups really do occur.

Less Conclusive Forms of Evidence

Readers should not take this discussion to imply that these are the only five forms of evidence regarding the potential genetic nature of between-group differences in intelligence. One early effort to address this issue was through transracial adoption studies where African American children were adopted into White families (e.g., Moore, 1986; Tizard, 1974). The most prominent and best designed of these studies was the Minnesota Transracial Adoption Study (MTAS), which started in the 1970s. In the first report (Scarr & Weinberg, 1976), African American and interracial (White and African American ancestry) children with a mean age of 6.2 years and adopted into White families at an average age of 18 months had modestly high IQ scores (M = 106.3, SD = 13.9). This average was above the national White IQ average (100) and only marginally lower than the mean IQ of White adoptees (111.5). These results seemed at the time to herald strong evidence against a genetic hypothesis for intergroup racial differences in intelligence. However, in the follow-up study, the adoptees' average IQs in adolescence had regressed toward their racial groups' mean. This was true for White adoptees (M = 105.6), interracial adoptees (M = 98.5), and African American adoptees (M = 89.4; Weinberg, Scarr, & Waldman, 1992). Thus, the evidence from the MTAS is ambiguous. Adoptees probably did receive an environmentally driven boost to their IQs, as indicated by the averages that were higher than each racial group's average IQ, which shows the importance of differences in home environments in average IQ gaps across racial groups in the United States. But the relative gap between African Americans' and White Americans' average IQs was still approximately 1 SD (16.2 IQ points in adolescence), and interracial children had average scores in between, all results in accordance with a model based on genetic influences. But these score differences could reflect the different preadoption experiences of the different groups. (For example, the African American adoptees were adopted later and had more preadoption placements than the other groups.) In a later discussion of the MTAS, the researchers acknowledged that the results did not permit strong inferences about the impact of either genetic or environmental influences (Waldman, Weinberg, & Scarr, 1994). I concur with this assessment.

With the best transracial adoption study producing ambiguous evidence (and having methodological shortcomings that, in retrospect, weakened any possible conclusions), it is clear that the current transracial adoption literature is not productive to analyze when discussing the likelihood of genetic influences on average differences in intelligence within the United States. The MTAS and other transracial adoption studies were noble efforts, but because they lack methodological characteristics necessary to demonstrate unconfounded environmental or genetic effects, their evidence is ambiguous. Their most problematic shortcoming is an absence of intelligence test scores for the adopted children's biological parents. A transracial adoption study with intelligence test scores for biological and adopted parents, children from every large American minority group, and data on the environment created by both sets of parents (including the prenatal environment) would provide direct evidence that is stronger than the data that I have presented in this article. Unfortunately, no such study currently exists, and there have been no reported American transracial adoption studies on intelligence since the MTAS. There are some studies of transracial adopted children published in other countries, but these studies usually involve international adoptions. None report intelligence test scores for biological parents of adopted children, and few have detailed data about the preadoption environment (see van IJzendoorn, Juffer, & Klein Poelhuis, 2005, for a review).

Another piece of evidence that is sometimes used to argue that average intelligence differences are not fully environmentally caused is the data on brain size (e.g., Rushton & Jensen, 2005). Although it is true that brain volume correlates positively with IQ and that East Asians have larger brains than Europeans, who have larger brains than Africans (all these differences being in averages, with large amounts of overlap between individuals in various groups), I do not find this evidence convincing for three reasons. First, biological differences are not necessarily genetic in origin (Gottfredson, 2009), and it has not been conclusively shown that interracial differences in mean brain size have a genetic cause *and* that this genetic cause also has a causal impact on the correlation between brain size and IQ. Second, the advantage for individuals with larger brains is not apparent when comparing male and female subjects; male brains are larger than female brains, even when body size is controlled for, but both sexes have the same average intelligence (Jensen, 1998; Jensen & Johnson, 1994). This finding is problematic for the argument that big brains are smarter brains. Finally, evidence is mounting that brain organization and functioning may be more important biological determinants of intelligence than sheer brain size (Haier, 2017).

Other sources of evidence (see Rushton & Jensen, 2005, for a summary) are ambiguous in their support for either the fully environmental or partially genetic theories of the causes of average group differences in intelligence. This is why I have limited myself to the five sources of evidence I discuss for the bulk of this article.

DISCUSSION

In this article I have presented five lines of indirect and direct evidence that mean group differences in intelligence in the United States are not 100% environmentally caused: findings in support of Spearman's hypothesis, consistent results from tests of measurement invariance across American racial groups, the mathematical relationship for between-group and within-group sources of heritability, the preliminary results of applying polygenic scores and the results of GWASs to non-European groups, and the results of admixture studies. No one line of evidence is incontrovertible proof that between-group differences in intelligence in the United States are partially genetic in origin (although the direct evidence is strongest). However, all five types of evidence point in the same direction: Racial group differences in intelligence are at least partially genetically caused. Meanwhile, there is no fully environmental theory that parsimoniously explains all the facts I have presented. Those wanting to adhere to a fully environmental explanation have so far constructed a series of ad hoc explanations, which lack the parsimony of a partially genetic explanation for the phenomena I describe (Gottfredson, 2005; Rushton & Jensen, 2005).

Moreover, it is not easy to see how current theories of fully environmental differences will accommodate the new direct evidence that is likely to emerge from the GWASs or future admixture studies, given the results that these methods have produced so far.

On the other hand, a partially genetic explanation not only can accommodate such findings but can generate new, testable predictions about the relationship between genes, environment, and intelligence phenotypes across populations. For example, my belief that polygenic scores from diverse GWAS samples will explain more phenotype variance in intelligence and be generalizable to non-European samples is strongly dependent on the assumption that between-group differences in intelligence are at least partially genetic. If this hypothesis is disproved, then it would be a major setback to the claim of a genetic influence on intergroup intelligence differences. The same is true of my (Warne, 2020a) hypothesis regarding intelligence being positively correlated with East Asian heritage in European–Asian admixed groups.

Despite the diverse psychometric and genetic data I have presented, one nagging question remains unanswered: What percentage of between-group variance in intelligence scores is due to genetic variance between racial and ethnic groups in the United States? None of the evidence I have presented in this article provides a conclusive answer to this question. The percentage is almost certainly greater than zero, but beyond that, these sources of information do not point to an exact estimate for between-group heritability. There is no scientific reason why this question remains unanswered in the 21st century. The methods and technology exist now to determine the between-group heritability for racial groups in the United States. I believe there are at least two reasons for this gap in knowledge: a remarkable lack of curiosity among scientists and the lack of funding to answer such a question.

Studies of racial differences in economic, health, educational, behavioral, and other outcomes are ubiquitous in the social sciences. That discrepancies in many outcomes exist is beyond doubt, and there is no shortage of hypothesized environmental causes. Yet researchers who study these differences rarely postulate genetic causes. Even when genetic causes of discrepancies are raised, some scientists concentrate on denying such possibilities instead of confronting the data and testing the hypotheses surrounding potential genetic causes for interracial group differences in intelligence (e.g., Kwate, 2001; Lewontin, 1970; Sternberg, 2005). For a group that is supposed to be curious about the world, the number of scientists forming and testing hypotheses about the possibility of a genetic influence on racial differences in intelligence is surprisingly small. One reason may be the lack of ideological diversity in the social sciences, with overwhelmingly left-of-center political beliefs among researchers (Inbar & Lammers, 2012; Kaufman, 2021; Redding, 2001) that may cause some social scientists to not investigate the genetics of group differences because they find the topic distasteful (e.g., Horgan, 2013; Sternberg, 2005) or believe they already know the answer (see Flynn, 2012, p. 36, for a discussion of this latter reason). If ideological conformity is discouraging research on average race differences in intelligence, then it is an impediment to scholarly progress because it prevents the creation of knowledge that emerges from the Hegelian clash of contradicting viewpoints—the thesis and antithesis—to arrive at a more correct understanding of facts: the synthesis.

Another impediment to determining the degree to which genetic influences affect intergroup differences in intelligence is the lack of funding. Government agencies and private foundations shy away from funding controversial topics, and the genetics of racial differences in intelligence is certainly one of those topics. In one recent case, recounted by mathematics researcher Theodore P. Hill (2018), he and Sergei Tabachnikov submitted an article on variability differences in intelligence between sexes as a possible explanation for the preponderance of men in the highest and lowest levels of academic accomplishment. Much to Hill's surprise, the National Science Foundation (NSF) asked the authors to remove the acknowledgement of Hill's NSF funding from the article, ostensibly because the topic was unrelated to Tabachnikov's funding. However, emails that Hill obtained from a Freedom of Information Act request showed that the real reason was pressure on the NSF from the chair of the Climate and Diversity Committee and the associate head for diversity and equity at Tabachnikov's university (T. P. Hill, 2018). If a government agency bowed to such political pressure regarding a simple acknowledgment of funding on an article about a topic that is uncontroversial among intelligence researchers,[3] the pressure to not award grants to researchers to study race differences in intelligence must be overwhelming. For some granting organizations, such an award may even be unthink-

able. The excuse that there are not sufficient funds for such research is unlikely. As James Flynn—no friend to the hypothesis of genetic causes of interracial differences in intelligence—stated,

> If universities have their way, the necessary research will never be done. They fund the most mundane research projects, but never seem to have funds to test for genetic differences between races. I tell US academics I can only assume that they believe that racial IQ differences have a genetic component, and fear what they might find. They never admit that the politics of race affects their research priorities. It is always just far more important to establish whether squirrels enjoy *The Magic Flute*. (Flynn, 2012, p. 36)

The aversion to investigating, funding, and promoting research on the genetic influences on average intelligence across racial groups has made the topic languish and has limited the amount of data available. As a result, the controversy has festered far longer than necessary, and it is difficult to arrive at exact conclusions based on strong empirical data. A full exploration of the causes of this dynamic (which are probably rooted in the sociology of science) is beyond the scope of this article. However, some scholars have postulated about causes of this phenomenon. Coleman (1991) suggested that a concern throughout the social sciences toward historically marginalized groups has led to an ethic of "conspicuous benevolence," where findings that reflect unfavorably on these groups go quietly unreported and studies that might reveal such findings are not conducted. Coleman believed that this ethic would lead to self-suppression of scholarly research in order for scientists to retain the approval of their colleagues, which is important for obtaining tenure, a promotion, favorable ratings for grant applications, or professional eminence.

A recent survey supports Coleman's (1991) view, finding that 43% of American academics and 30% of British academics in the humanities and social sciences stated that they would support the firing of a colleague who published work showing that hiring more minorities and women is correlated with lower organizational performance (Kaufman, 2021, p. 24). This scenario engendered more opposition than any of the other four controversial topics in the survey. A different survey of undergraduate students found that 85% of respondents would oppose a speaker on campus who claimed that "Some racial groups are less intelligent than others" (College Pulse, Foundation for Individual Rights in Education, & RealClear Education, n.d., p. 55). These surveys shows that Coleman (1991) may have been too mild; academics may not be avoiding race differences in IQ solely to retain approval from their colleagues, as important as that may be. They may be trying to keep their jobs or avoid being the center of a firestorm. These strong negative reactions from students and faculty seem to occur despite the data that indicate that a partially genetic cause of group differences in IQ is plausible.

Other reasons for the aversion toward researching group differences in IQ have been suggested. Davis (1978) believed that a "moralistic fallacy" was at work, where people's moral desires for what they want to be true stand in place for actual scientific truth—a cognitive heuristic that leads people to make decisions about scientific discovery accordingly. Humphreys (1988, 1991) suggested several widespread folk beliefs among social scientists that would discourage research on the possibility of a genetic influence on mean intelligence differences, including a bias among social scientists in favor of environmental causes of human behavior, a narrow focus on socioeconomic status as a powerful cause of individual and group differences in life outcomes, and a default belief that there could not be any genetic influences on group differences. Regardless of the cause, the result is that the topic of genetic influences on intergroup intelligence averages has been understudied, which allows the controversy to linger on unnecessarily for decades and prevents social scientists from crafting policy proposals based on sound data.

CONCLUSION

The five types of circumstantial evidence presented here are not, individually, conclusive regarding the potential of genetic influences on mean differences among racial groups in the United States. Each has its own strengths and weaknesses. For example, the GWAS evidence consists of just three studies, but these studies provide consistent direct evidence of the influence of specific genes on intelligence in different racial groups. The measurement invariance evidence cannot provide an estimate of the degree

to which between-group variance is genetic, but it does eliminate the possibility of a unique influence that uniformly lowers one group's intelligence test scores while leaving another group untouched (see Warne, 2020b, Chapters 29–30, for a further discussion of the unlikelihood of such unique influences). Combined, all five sources of data make a cohesive argument that group differences in intelligence are not fully environmentally caused.

I call upon psychologists to have an open mind and to investigate the evidence for themselves, starting with the sources I have cited in this article. I also encourage social scientists to make research contributions that can address this question. If readers believe in fully environmental causes for mean racial group differences in intelligence, I invite them to design their own studies to test their hypothesis and to create a coherent theory of environmental causes of intelligence that can parsimoniously explain the findings I have summarized and make new, testable predictions. Controversy thrives in ignorance; it is time to move beyond the basic question of whether genes have an impact on intergroup mean IQs so that scientists can investigate productive questions of the magnitude of this influence and how to accommodate differences in a diverse society.

NOTES

I have no conflicts of interest to disclose.

Address correspondence about this article to Russell T. Warne, 800 W. University Parkway MC 115, Orem, UT 84058 (email: rwarne@uvu.edu).

1. Incidentally, this discussion demonstrates the power of random assignment in true experiments. Random assignment of subjects to experimental and control groups balances out these groups genetically, which means that any later differences on the dependent variable(s) cannot be due to genetic differences between groups. Given the ubiquity of genetic influences on behavior (Plomin et al., 2016), random assignment is a powerful protection against genetic confounding.

2. This estimate is derived by squaring the Cohen's *d* values in Table 2 and dividing them by the squared σ_E values in Table 1 for an h_b^2 value of zero and an h_w^2 value between .20 and .80. The resulting values of 7.3% and 47.6% are the minimum and maximum percentages of needed total environmental variance shared with socioeconomic status variance that the Cohen's *d* values in Table 2 could produce. As would be expected mathematically, higher percentages result from larger socioeconomic status differences in Table 2 and lower h_w^2 values in Table 1.

3. See Feingold (1992) for a landmark literature review on the topic. The finding of greater male variability is now widely accepted, with many different tests usually showing a 5–15% greater standard deviation on male cognitive test scores (e.g., Deary, Thorpe, Wilson, Starr, & Whalley, 2003; Hur, te Nijenhuis, & Jeong, 2017; Lakin, 2013; Strand, Deary, & Smith, 2006).

REFERENCES

1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. *Nature, 526*(7571), 68–74. https://doi.org/10.1038/nature15393

Adams, H. H. H., Hibar, D. P., Chouraki, V., Stein, J. L., Nyquist, P. A., Rentería, M. E., . . . Thompson, P. M. (2016). Novel genetic loci underlying human intracranial volume identified through genome-wide association. *Nature Neuroscience, 19,* 1569–1583. https://doi.org/10.1038/nn.4398

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Arden, R., Luciano, M., Deary, I. J., Reynolds, C. A., Pedersen, N. L., Plassman, B. L., . . . Visscher, P. M. (2016). The association between intelligence and lifespan is mostly genetic. *International Journal of Epidemiology, 45,* 178–185. https://doi.org/10.1093/ije/dyv112

Asbury, K., Moran, N., & Plomin, R. (2016). Nonshared environmental influences on academic achievement at age 16: A qualitative hypothesis-generating monozygotic-twin differences study. *AERA Open, 2*(4). https://doi.org/10.1177/2332858416673596

Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence, 33,* 431–444. https://doi.org/10.1016/j.intell.2004.12.004

Beaujean, A. A., McGlaughlin, S. M., & Margulies, A. S. (2009). Factorial validity of the Reynolds Intellectual Assessment Scales for referred students. *Psychology in the Schools, 46,* 932–950. https://doi.org/10.1002/pits.20435

Binet, A., & Simon, T. (1905/1916). New methods for the diagnosis of the intellectual level of subnormals (E. S. Kite, trans.). In A. Binet & T. Simon, *The development of intelligence in children (the Binet–Simon Scale)* (pp. 37–90). Williams & Wilkins.

Bouchard, T. Jr. (2014). Genes, evolution and intelligence. *Behavior Genetics, 44,* 549–577. https://doi.org/10.1007/s10519-014-9646-x

Bouchard, T. J. Jr., Lykken, D. T., Tellegen, A., & McGue, M. (1996). Genes, drives, environment, and experience: EPD theory revised. In C. P. Benbow (Ed.), *Intellectual talent: Psychometric and social issues* (pp. 5–43). Johns Hopkins University Press.

Bryc, K., Durand, E. Y., Macpherson, J. M., Reich, D., & Mountain, J. L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics, 96,* 37–53. https://doi.org/10.1016/j.ajhg.2014.11.010

Burks, B. S. (1973). 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. In H. J. Eysenck (Ed.), *The measurement of intelligence* (pp. 325–357). Springer. (Original work published in 1928)

Cheng, C.-Y., Reich, D., Haiman, C. A., Tandon, A., Patterson, N., Selvin, E., . . . Kao, W. H. L. (2012). African ancestry and its correlation to type 2 diabetes in African Americans: A genetic admixture analysis in three U.S. Population cohorts. *PLOS ONE, 7*(3), e32840. https://doi.org/10.1371/journal.pone.0032840

Cofnas, N. (2016). Science is not always "self-correcting": Fact–value conflation and the study of intelligence. *Foundations of Science, 21*(3), 477–492. https://doi.org/10.1007/s10699-015-9421-3

Coleman, J. S. (1991). The Sidney Hook Memorial Award address: On the self-suppression of academic freedom. *Academic Questions, 4*(1), 17–22. https://doi.org/10.1007/BF02682943

College Pulse, Foundation for Individual Rights in Education, & RealClear Education. (n.d.). *2020 College Free Speech Rankings: What's the climate for free speech on America's college campuses?* https://www.thefire.org/research/publications/student-surveys/2020-college-free-speech-rankings/

Curran, F. C., & Kellogg, A. T. (2016). Understanding science achievement gaps by race/ethnicity and gender in kindergarten and first grade. *Educational Researcher, 45,* 273–282. https://doi.org/10.3102/0013189x16656611

Dahlke, J. A., & Sackett, P. R. (2017). The relationship between cognitive-ability saturation and subgroup mean differences across predictors of job performance. *Journal of Applied Psychology, 102,* 1403–1420. https://doi.org/10.1037/apl0000234

Dalliard, M. (2014). The elusive x-factor: A critique of J. M. Kaplan's model of race and IQ. *Open Differential Psychology*. https://doi.org/10.26775/ODP.2014.08.25

Damian, R. I., Su, R., Shanahan, M., Trautwein, U., & Roberts, B. W. (2015). Can personality traits and intelligence compensate for background disadvantage? Predicting status attainment in adulthood. *Journal of Personality and Social Psychology, 109,* 473–489. https://doi.org/10.1037/pspp0000024

Davis, B. B. (1978). The moralistic fallacy. *Nature, 272*(5652), 390. https://doi.org/10.1038/272390a0

Deary, I. J. (2012). 125 years of intelligence in *The American Journal of Psychology*. *American Journal of Psychology, 125,* 145–154. https://doi.org/10.5406/amerjpsyc.125.2.0145

Deary, I. J., Penke, L., & Johnson, W. (2010). The neuroscience of human intelligence differences. *Nature Reviews: Neuroscience, 11,* 201–211. https://doi.org/10.1038/nrn2793

Deary, I. J., Thorpe, G., Wilson, V., Starr, J. M., & Whalley, L. J. (2003). Population sex differences in IQ at age 11: The Scottish mental survey 1932. *Intelligence, 31,* 533–542. https://doi.org/10.1016/s0160-2896(03)00053-9

DeFries, J. C. (1972). Quantitative aspects of genetics and environment in the determination of behavior. In L. Ehrman, G. S. Omenn, & E. Caspari (Eds.), *Genetics, environment, and behavior* (pp. 5–16). Academic Press.

Derrick, S. M. (1920). A comparative study of the intelligence of seventy-five whites and fifty-five colored college students by the Stanford revision of the Binet–Simon scale. *Journal of Applied Psychology, 4,* 316–329. https://doi.org/10.1037/h0071332

Detterman, D. K. (2006). Editorial note on controversial papers. *Intelligence, 34,* iv. https://doi.org/10.1016/j.intell.2005.10.001

Dickens, W. T., & Flynn, J. R. (2006). Black Americans reduce the racial IQ gap: Evidence from standardization samples. *Psychological Science, 17,* 913–920. https://doi.org/10.1111/j.1467-9280.2006.01802.x

Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research, 35,* 21–50. https://doi.org/10.1207/S15327906MBR3501_2

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 F. H. Columbus (Ed.), *Advances in psychology research* (Vol. 6, pp. 31–59). Nova Science Publishers.

Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M., & Boardman, J. D. (2015). Polygenic influence on educational attainment: New evidence from the National Longitudinal Study of Adolescent to Adult Health. *AERA Open, 1*(3), 1–13. https://doi.org/10.1177/2332858415599972

Dunkel, C. S., Woodley of Menie, M. A., Pallesen, J., & Kirkegaard, E. O. W. (2019). Polygenic scores mediate the Jewish phenotypic advantage in educational attainment and cognitive ability compared with Catholics and Lutherans. *Evolutionary Behavioral Sciences, 13,* 366–375. https://doi.org/10.1037/ebs0000158

Feingold, A. (1992). Sex differences in variability in intellectual abilities: A new look at an old controversy. *Review of Educational Research, 62,* 61–84. https://doi.org/10.3102/00346543062001061

Flashman, L. A., Andreasen, N. C., Flaum, M., & Swayze, V. W. II. (1997). Intelligence and regional brain volumes

in normal controls. *Intelligence, 25,* 149–160. https://doi.org/10.1016/S0160-2896(97)90039-8

Flores, C., Ma, S.-F., Pino-Yanes, M., Wade, M. S., Pérez-Méndez, L., Kittles, R. A., . . . Garcia, J. G. N. (2012). African Ancestry is associated with asthma risk in African Americans. *PLOS ONE, 7*(1), e26807. https://doi.org/10.1371/journal.pone.0026807

Flynn, J. R. (1980). *Race, IQ, and Jensen*. Routledge & Kegan Paul.

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

Flynn, J. R., te Nijenhuis, J., & Metzen, D. (2014). The *g* beyond Spearman's *g*: Flynn's paradoxes resolved using four exploratory meta-analyses. *Intelligence, 44,* 1–10. https://doi.org/10.1016/j.intell.2014.01.009

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. https://doi.org/10.1016/j.intell.2015.04.007

Genç, E., Fraenz, C., Schlüter, C., Friedrich, P., Hossiep, R., Voelkle, M. C., . . . Jung, R. E. (2018). Diffusion markers of dendritic density and arborization in gray matter predict differences in intelligence. *Nature Communications, 9,* 1905. https://doi.org/10.1038/s41467-018-04268-8

Gignac, G. E., & Bates, T. C. (2017). Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence, 64,* 18–29. https://doi.org/10.1016/j.intell.2017.06.004

Gottfredson, L. S. (1997a). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography. *Intelligence, 24,* 13–23. https://doi.org/10.1016/S0160-2896(97)90011-8

Gottfredson, L. S. (1997b). Why *g* matters: The complexity of everyday life. *Intelligence, 24,* 79–132. https://doi.org/10.1016/S0160-2896(97)90014-3

Gottfredson, L. S. (2005). What if the hereditarian hypothesis is true? *Psychology, Public Policy, and Law, 11,* 311–319. https://doi.org/10.1037/1076-8971.11.2.311

Gottfredson, L. S. (2009). Logical fallacies used to dismiss the evidence on intelligence testing. In R. P. Phelps (Ed.), *Correcting fallacies about educational and psychological testing* (pp. 11–65). American Psychological Association.

Gottfredson, L. S., & Deary, I. J. (2004). Intelligence predicts health and longevity, but why? *Current Directions in Psychological Science, 13,* 1–4. https://doi.org/10.1111/j.0963-7214.2004.01301001.x

Gould, S. J. (1981). *The mismeasure of man*. W. W. Norton.

Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. *Multivariate Behavioral Research, 27,* 175–204. https://doi.org/10.1207/s15327906mbr2702_2

Haier, R. J. (2017). *The neuroscience of intelligence*. Cambridge University Press.

Hartmann, P., Kruuse, N. H. S., & Nyborg, H. (2007). Testing the cross-racial generality of Spearman's hypothesis in two samples. *Intelligence, 35,* 47–57. https://doi.org/10.1016/j.intell.2006.04.004

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47,* 1083–1101. https://doi.org/10.1037/0003-066X.47.9.1083

Herrnstein, R. J., & Murray, C. (1996). *The bell curve: Intelligence and class structure in American life* (2nd ed.). Free Press.

Hill, T. P. (2018, September 7). Academic activists send a published paper down the memory hole. *Quillette*. https://quillette.com/2018/09/07/academic-activists-send-a-published-paper-down-the-memory-hole/

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., . . . Deary, I. J. (2019). A combined analysis of genetically correlated traits identifies 187 loci and a role for neurogenesis and myelination in intelligence. *Molecular Psychiatry, 24,* 169–181. https://doi.org/10.1038/s41380-017-0001-5

Horgan, J., (2013, May 16). Should research on race and IQ be banned? *Scientific American Cross-Check Blog*. https://blogs.scientificamerican.com/cross-check/should-research-on-race-and-iq-be-banned/

Hu, M., Lasker, J., Kirkegaard, E. O. W., & Fuerst, J. G. R. (2019). Filling in the gaps: The association between intelligence and both color and parent-reported ancestry in the National Longitudinal Survey of Youth 1997. *Psych, 1,* 240–261. https://doi.org/10.3390/psych1010017

Humphreys, L. G. (1988). Trends in levels of academic achievement of blacks and other minorities. *Intelligence, 12,* 231–260. https://doi.org/10.1016/0160-2896(88)90025-6

Humphreys, L. G. (1991). Limited vision in the social sciences. *American Journal of Psychology, 104, 333–353*. https://doi.org/10.2307/1423243

Hunt, E. (2011). *Human intelligence*. Cambridge University Press.

Hur, Y.-M., te Nijenhuis, J., & Jeong, H.-U. (2017). Testing Lynn's theory of sex differences in intelligence in a large sample of Nigerian school-aged children and adolescents (n > 11,000) using Raven's Standard Progressive Matrices Plus. *Mankind Quarterly, 57,* 428–437.

Inbar, Y., & Lammers, J. (2012). Political diversity in social and personality psychology. *Perspectives on Psychological Science, 7,* 496–503. https://doi.org/10.1177/1745691612448792

Jencks, C. (1979). *Who gets ahead? The determinants of economic success in America*. Basic Books.

Jensen, A. R. (1980a). *Bias in mental testing*. The Free Press.

Jensen, A. R. (1980b). Précis of bias in mental testing. *Behavioral and Brain Sciences, 3,* 325–333. https://doi.org/10.1017/S0140525X00005161

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. https://doi.org/10.1017/S0140525X00020392

Jensen, A. R. (1992). Spearman's hypothesis: Methodology and evidence. *Multivariate Behavioral Research, 27,* 225–233. https://doi.org/10.1207/s15327906mbr2702_5

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

Jensen, A. R., & Johnson, F. W. (1994). Race and sex differences in head size and IQ. *Intelligence, 18,* 309–333. https://doi.org/10.1016/0160-2896(94)90032-9

Jin, W., Xu, S., Wang, H., Yu, Y., Shen, Y., Wu, B., & Jin, L. (2012). Genome-wide detection of natural selection in African Americans pre- and post-admixture. *Genome Research, 22,* 519–527. https://doi.org/10.1101/gr.124784.111

Jung, R. E., & Haier, R. J. (2007). The parieto-frontal integration theory (P-FIT) of intelligence: Converging neuroimaging evidence. *Behavioral and Brain Sciences, 30,* 135–154. https://doi.org/10.1017/S0140525X07001185

Kane, H. (2007). Group differences in nonverbal intelligence: Support for the influence of Spearman's *g*. *Mankind Quarterly, 48,* 65–82.

Kao, W. H. L., Klag, M. J., Meoni, L. A., Reich, D., Berthier-Schaad, Y., Li, M., . . . Parekh, R. S. (2008). MYH9 is associated with nondiabetic end-stage renal disease in African Americans. *Nature Genetics, 40,* 1185–1192. https://doi.org/10.1038/ng.232

Kaplan, J. M. (2015). Race, IQ, and the search for statistical signals associated with so-called "X"-factors: Environments, racism, and the "hereditarian hypothesis." *Biology & Philosophy, 30,* 1–17. https://doi.org/10.1007/s10539-014-9428-0

Kaufman, E. (2021, March 1). *Academic freedom in crisis: Punishment, political discrimination, and self-censorship* (Report no. 2). Center for the Study of Partisanship and Ideology. https://cspicenter.org/wp-content/uploads/2021/03/AcademicFreedom.pdf

Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: A systematic review. *Psychological Medicine, 37,* 615–626. https://doi.org/10.1017/S0033291706009524

Kirkegaard, E. O. W., Woodley of Menie, M. A., Williams, R. L., Fuerst, J., & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych, 1,* 1–25. https://doi.org/10.3390/Psychology1010001

Kourany, J. A. (2016). Should some knowledge be forbidden? The case of cognitive differences research. *Philosophy of Science, 83,* 779–790. https://doi.org/https://doi.org/10.1086/687863

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J.-B., . . . Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences, 111,* 15273–15278. https://doi.org/10.1073/pnas.1408777111

Krieger, N., Sidney, S., & Coakley, E. (1998). Racial discrimination and skin color in the CARDIA study: Implications for public health research. *American Journal of Public Health, 88,* 1308–1313. https://doi.org/10.2105/ajph.88.9.1308

Kwate, N. O. A. (2001). Intelligence or misorientation? Eurocentrism in the WISC-III. *Journal of Black Psychology, 27,* 221–238. https://doi.org/10.1177/0095798401027002005

Lakin, J. M. (2013). Sex differences in reasoning abilities: Surprising evidence that male–female ratios in the tails of the quantitative reasoning distribution have increased. *Intelligence, 41,* 263–274. https://doi.org/10.1016/j.intell.2013.04.004

Lasker, J., Pesta, B. J., Fuerst, J. G. R., & Kirkegaard, E. O. W. (2019). Global ancestry and cognitive ability. *Psych, 1,* 431–459. https://doi.org/10.3390/psych1010034

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., . . . Cesarini, D. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics, 50*(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Lewontin, R. C. (1970). Race and intelligence. *Bulletin of the Atomic Scientists, 26*(3), 2–8. https://doi.org/10.1080/00963402.1970.11457774

Li, X., Sano, H., & Merwin, J. C. (1996). Perception and reasoning abilities among American, Japanese, and Chinese adolescents. *Journal of Adolescent Research, 11,* 173–193. https://doi.org/10.1177/0743554896112002

Lipsey, M. W., Farran, D. C., & Durkin, K. (2018). Effects of the Tennessee Prekindergarten Program on children's achievement and behavior through third grade. *Early Childhood Research Quarterly, 45,* 155–176. https://doi.org/10.1016/j.ecresq.2018.03.005

Little, M. (2017). Racial and socioeconomic gaps in executive function skills in early elementary school. *Educational Researcher, 46,* 103–109. https://doi.org/10.3102/0013189x17698700

Lubke, G. H., Dolan, C. V., & Kelderman, H. (2001). Investigating group differences on cognitive tests using Spearman's hypothesis: An evaluation of Jensen's method. *Multivariate Behavioral Research, 36,* 299–324. https://doi.org/10.1207/s15327906299-324

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. https://doi.org/10.1016/S0160-2896(03)00051-5

Maller, S. J. (2000). Item invariance in four subtests of the Universal Nonverbal Intelligence Test (UNIT) across groups of deaf and hearing children. *Journal of Psychoeducational Assessment, 18,* 240–254. https://doi.org/10.1177/073428290001800304

Martin, A. R., Gignoux, C. R., Walters, R. K., Wojcik, G. L., Neale, B. M., Gravel, S., . . . Kenny, E. E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *American Journal of Human Genetics, 100,* 635–649. https://doi.org/10.1016/j.ajhg.2017.03.004

Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika, 58,* 525–543. https://doi.org/10.1007/BF02294825

Mollon, J., Knowles, E. E. M., Mathias, S. R., Gur, R., Peralta, J. M., Weiner, D. J., . . . Glahn, D. C. (2021). Genetic influence on cognitive development between childhood and adulthood. *Molecular Psychiatry, 26,* 656–665. https://doi.org/10.1038/s41380-018-0277-0

Moore, E. G. (1986). Family socialization and the IQ test performance of traditionally and transracially adopted Black children. *Developmental Psychology, 22,* 317–326. https://doi.org/10.1037/0012-1649.22.3.317

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. https://doi.org/10.1016/j.intell.2007.02.001

Murray, C. (2020). *Human diversity: The biology of gender, race, and class*. Twelve.

Murray, C. (2021). *Facing reality: Two truths about race in America*. Encounter Books.

Nagoshi, C. T., Johnson, R. C., DeFries, J. C., Wilson, J. R., & Vandenberg, S. G. (1984). Group differences and first principal-component loadings in the Hawaii family study of cognition: A test of the generality of "Spearman's hypothesis." *Personality and Individual Differences, 5,* 751–753. https://doi.org/10.1016/0191-8869(84)90125-9

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., . . . Urbina, S. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101. https://doi.org/10.1037/0003-066X.51.2.77

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012a). Group differences in IQ are best understood as environmental in origin. *American Psychologist, 67*(6), 503–504. https://doi.org/10.1037/a0029772

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012b). Intelligence: New findings and theoretical developments. *American Psychologist, 67,* 130–159. https://doi.org/10.1037/a0026699

Ogbu, J. U. (2002). Cultural amplifiers of intelligence: IQ and minority status in cross-cultural perspective. In J. M. Fish (Ed.), *Race and intelligence: Separating science from myth* (pp. 241–278). Erlbaum.

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature, 533*(7604), 539–542. https://doi.org/10.1038/nature17671

Penke, L., Maniega, S. M., Bastin, M. E., Valdes Hernandez, M. C., Murray, C., Royle, N. A., . . . Deary, I. J. (2012). Brain white matter tract integrity as a neural foundation for general intelligence. *Molecular Psychiatry, 17,* 1026–1030. https://doi.org/10.1038/mp.2012.66

Pesta, B. J., Kirkegaard, E. O. W., te Nijenhuis, J., Lasker, J., & Fuerst, J. G. R. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence, 78,* 101408. https://doi.org/10.1016/j.intell.2019.101408

Piffer, D. (2015). A review of intelligence GWAS hits: Their relationship to country IQ and the issue of spatial autocorrelation. *Intelligence, 53,* 43–50. https://doi.org/10.1016/j.intell.2015.08.008

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. https://doi.org/10.3390/psych1010005

Plomin, R. (2018). *Blueprint: How DNA makes us who we are*. MIT Press.

Plomin, R., & Deary, I. J. (2015). Genetics and intelligence differences: Five special findings. *Molecular Psychiatry, 20,* 98–108. https://doi.org/10.1038/mp.2014.105

Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science, 11,* 3–23. https://doi.org/10.1177/1745691615617439

Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics, 19*(3), 148–159. https://doi.org/10.1038/nrg.2017.104

Popejoy, A. B., & Fullerton, S. M. (2016). Genomics is failing on diversity. *Nature, 538*(7624), 161–164. https://doi.org/10.1038/538161a

Redding, R. E. (2001). Sociopolitical diversity in psychology: The case for pluralism. *American Psychologist, 56,* 205–215. https://doi.org/10.1037/0003-066X.56.3.205

Reich, D. (2018). *Who we are and how we got here: Ancient DNA and the new science of the human past*. Pantheon Books.

Reynolds, C. R. (2000). Why is psychometric research on bias in mental testing so often ignored? *Psychology, Public Policy, and Law, 6,* 144–150. https://doi.org/10.1037/1076-8971.6.1.144

Rushton, J. P. (1999). Secular gains in IQ not related to the *g* factor and inbreeding depression—unlike Black–White differences: A reply to Flynn. *Personality and Individual Differences, 26,* 381–389. https://doi.org/10.1016/S0191-8869(98)00148-2

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11,* 235–294. https://doi.org/10.1037/1076-8971.11.2.235

Rushton, J. P., & Jensen, A. R. (2010). The rise and fall of the Flynn effect as a reason to expect a narrowing of the Black–White IQ gap. *Intelligence, 38,* 213–219. https://doi.org/10.1016/j.intell.2009.12.002

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., de Leeuw, C. A., . . . Posthuma, D. (2018). Genome-wide association meta-analysis in 269,867 individuals identifies new genetic and functional links to intelligence. *Nature Genetics, 50,* 912–919. https://doi.org/10.1038/s41588-018-0152-6

Scarr, S., & Weinberg, R. A. (1976). IQ test performance of Black children adopted by White families. *American Psychologist, 31,* 726–739. https://doi.org/10.1037/0003-066x.31.10.726

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15,* 201–292. https://doi.org/10.2307/1412107

Spearman, C. (1927). *The abilities of man: Their nature and measurement*. Macmillan.

Sternberg, R. J. (2005). There are no public-policy implications: A reply to Rushton and Jensen (2005). *Psychology, Public Policy, and Law, 11,* 295–301. https://doi.org/10.1037/1076-8971.11.2.295

Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in Cognitive Abilities Test scores: A UK national picture. *British Journal of Educational Psychology, 76,* 463–480. https://doi.org/10.1348/000709905x50906

Tang, H., Quertermous, T., Rodriguez, B., Kardia, S. L. R., Zhu, X., Brown, A., . . . Risch, N. J. (2005). Genetic structure, self-identified race/ethnicity, and confounding in case-control association studies. *American Journal of Human Genetics, 76*(2), 268–275. https://doi.org/10.1086/427888

te Nijenhuis, J., Jongeneel-Grimen, B., & Armstrong, E. L. (2015). Are adoption gains on the *g* factor? A meta-analysis. *Personality and Individual Differences, 73,* 56–60. https://doi.org/10.1016/j.paid.2014.09.022

te Nijenhuis, J., & van den Hoek, M. (2016). Spearman's hypothesis tested on black adults: A meta-analysis. *Journal of Intelligence, 4,* 6. https://doi.org/10.3390/jintelligence4020006

te Nijenhuis, J., van den Hoek, M., & Armstrong, E. L. (2015). Spearman's hypothesis and Amerindians: A meta-analysis. *Intelligence, 50,* 87–92. https://doi.org/10.1016/j.intell.2015.02.006

te Nijenhuis, J., van den Hoek, M., & Dragt, J. (2019). A meta-analysis of Spearman's hypothesis tested on Latin-American Hispanics, including a new way to correct for imperfectly measuring the construct of *g*. *Psych, 1,* 101–122. https://doi.org/10.3390/psych1010008

Tizard, B. (1974). IQ and race. *Nature, 247*(5439), 316. https://doi.org/10.1038/247316a0

Turkheimer, E. (2007). Race and IQ. *Cato Unbound*. https://www.cato-unbound.org/2007/11/21/eric-turkheimer/race-iq

U.S. Department of Health and Human Services. (2012). Third grade follow-up to *Head Start impact study. Final report* (OPRE report 2012-45).

van IJzendoorn, M. H., Juffer, F., & Klein Poelhuis, C. W. (2005). Adoption and cognitive development: A meta-analytic comparison of adopted and nonadopted children's IQ and school performance. *Psychological Bulletin, 131,* 301–316. https://doi.org/10.1037/0033-2909.131.2.301

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. https://doi.org/10.1016/0160–2896(94)90051-5

Warne, R. T. (2016a). Five reasons to put the *g* back into giftedness: An argument for applying the Cattell–Horn–Carroll theory of intelligence to gifted education research and practice. *Gifted Child Quarterly, 60,* 3–15. https://doi.org/10.1177/0016986215605360

Warne, R. T. (2016b). Testing Spearman's hypothesis with advanced placement examination data. *Intelligence, 57,* 87–95. https://doi.org/10.1016/j.intell.2016.05.002

Warne, R. T. (2020a). Continental genetic ancestry source correlates with global cognitive ability score. *Mankind Quarterly, 60,* 400–422.

Warne, R. T. (2020b). *In the know: Debunking 35 myths about human intelligence*. Cambridge University Press. https://doi.org/10.1017/9781108593298

Warne, R. T., Astle, M. C., & Hill, J. C. (2018). What do undergraduates learn about human intelligence? An analysis of introductory psychology textbooks. *Archives of Scientific Psychology, 6,* 32–50. https://doi.org/10.1037/arc0000038

Warne, R. T., Yoon, M., & Price, C. J. (2014). Exploring the various interpretations of "test bias." *Cultural Diversity & Ethnic Minority Psychology, 20,* 570–582. https://doi.org/10.1037/a0036503

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. https://doi.org/10.1016/0160-2896(92)90028-p

Wicherts, J. M. (2017). Psychometric problems with the method of correlated vectors applied to item scores (including some nonsensical results). *Intelligence, 60,* 26–38. https://doi.org/10.1016/j.intell.2016.11.002

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., & Span, M. M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence, 32,* 509–537. https://doi.org/10.1016/j.intell.2004.07.002

Woodley, M. A., te Nijenhuis, J., Must, O., & Must, A. (2014). Controlling for increased guessing enhances the independence of the Flynn effect from *g*: The return of the Brand effect. *Intelligence, 43,* 27–34. https://doi.org/10.1016/j.intell.2013.12.004

Yang, J., Bakshi, A., Zhu, Z., Hemani, G., Vinkhuyzen, A. A. E., Lee, S. H., . . . Visscher, P. M. (2015). Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. *Nature Genetics, 47*(10), 1114–1120. https://doi.org/10.1038/ng.3390