Race, IQ, and the Search for Statistical Signals Associated with So-Called "X"-Factors: Environments, Racism, and the "Hereditarian Hypothesis"

Citation	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), 1-17. doi:10.1007/s10539-014-9428-0
DOI	10.1007/s10539-014-9428-0
Publisher	Springer
Version	Accepted Manuscript
Terms of Use	http://cdss.library.oregonstate.edu/sa-termsofuse

Title Page

Author: Jonathan Michael Kaplan

Title: Race, IQ, and the Search for Statistical Signals Associated with So-Called “X”-Factors: Environments, Racism, and the “Hereditarian Hypothesis”

Author Affiliation:
Philosophy
School of History, Philosophy, and Religion
Oregon State University
Corvallis, OR 97331
Email: jonathan.kaplan@oregonstate.edu
Phone: 541-224-2994

Abstract: Some authors defending the "hereditarian" hypothesis with respect to differences in average IQ scores between populations have argued that the sorts of environmental variation hypothesized by some researchers rejecting the hereditarian position should leave discoverable statistical traces, namely changes in the overall variance of scores or in variance-covariance matrixes relating scores to other variables. In this paper, I argue that the claims regarding the discoverability of such statistical signals are broadly mistaken - there is no good reason to suspect that the hypothesized environmental causes would leave detectable traces of the sorts suggested. As there remains no way to gather evidence that would permit the direct refutation of the environmental hypotheses, and no direct evidence for the hereditarian position, it remains the case, I argue, that the hereditarian position is unsupported by current evidence.

Keywords: Race; IQ; intelligence; hereditarian; environmental; variance; statistical signals

Acknowledgements: [Suppressed for anonymous review]

Race, IQ, and the Search for Statistical Signals Associated with So-Called "X"-Factors:

Environments, Racism, and the "Hereditarian Hypothesis"

I. Introduction: Race, "X Factors" and "Hereditarian Hypothesis"

In his defense of the possibility of treating "race" as a straightforwardly biological concept, Sesardic (2010) claimed that the so-called "hereditarian position"– that is, the attribution of "the observed IQ differences between racial groups" to, at least in part, "genetic difference between them" – "is actually supported by a number of empirical arguments that have considerable strength and legitimacy" (Sesardic 2010 157-158). This view is, I think, mistaken; the empirical arguments cited to support the hereditiarn hypothesis are neither particularly strong, nor should they be regarded in any obvious way as "legitimate." The reason for this is, I argue, straightforward: given the actual state of the world there is no way to generate any reasonably strong evidence in favor of the hereditarian hypothesis.

The hereditarian position, recall, is that the so-called "IQ gap" between Black and White Americans[1] – a difference of perhaps 1 standard deviation (that is, around 15

[1] Hereditarian hypotheses are sometimes extended to include the average differences in test scores between Blacks, Whites, and "Asians" (with "Asians" outperforming "Caucasians" by about a third of a standard deviation) (Herrnstein and Murray 1994; Rushton and Jensen 2005a), or the differences in test scores between smaller populations (Lynn and Vanhanen 2002). While the focus of this paper is on the claims

points, but varying somewhat depending on who is doing the measuring and under what conditions[2]) – is due at least in large part[3] to differences in the "races" average genetic endowments directly relevant to the development of the sorts of abilities tested on IQ tests (and related performance measures). The exact details of the proposed pathways between the average genetic differences between the so-called "races" and the average differences in performance on IQ tests (and related measures) are rarely explored in the literature, but, negatively, the hereditarian hypothesis demands that *racism* not be a major mediating factor. The pathway must be reasonably direct (e.g. not mediated by the differences in treatment that genes associated with e.g. skin color and other physical features engender), and the difference in realized performance must be robust across a reasonably wide range of developmental environments (so unlikely to change without more radical changes in developmental environments than usually envisioned).

---

surrounding Black/White differences, for the most part, the same analysis applies to the extended theories.

[2] Herrnstein and Murray (1994) claim about 1.2 standard deviations, or about 17 points, as the best estimate, but use a "low" estimate of 1 standard deviation or about 15 points in their arguments (277-279); Rushton and Jensen (2005a) cite 1.1 standard deviations as the best estimate. Dickens and Flynn (2006a) argue that the "gap" has been narrowing over time, and that estimates of the *current* gap should be substantially lower (perhaps 4-7 points lower); Rushton and Jensen suggest they deny this (2006), but see Dickens and Flynn's 2006b response. Nisbett (2005) argues that an estimate of .6 - .7 standard deviations, or about 10 points, is the most reasonable current estimate.

[3] Numbers between 50% and 80+% are often cited; see Rushton and Jensen 2005a.

Rather than attempt a complete review of the lines of arguments put forward in favor of (and, for that matter, opposed to) the hereditarian hypothesis, this paper argues for two related claims that together strongly imply that there is, given the actually state of the world, no way to generate good evidence in favor of the hereditarian hypothesis. First, with respect to the kinds of environmental variation associated with differences in performance on IQ tests (and related measured), the environments experienced by Black Americans (and Blacks more generally) are importantly different from the environments experienced by White Americans in ways directly attributable to race, and, more particularly, to racism. As discussed below, these kinds of environmental differences – differences not merely in the frequency with which particular kinds of environments are encountered, but in the kinds of environments themselves – are referred to in the literature as "X-factors," to distinguish them from the "ordinary" environmental differences that exist throughout the populations in question ("VE-factors"). Given that the environments encountered differ in these ways, attempts to statistically control for the effects of the different environments experienced by the members of the two populations are therefore, I argue, doomed to failure.

Second, I argue that searches for statistical signals for such racialized environmental differences ("X-factors") are unlikely to succeed, given reasonable assumptions about the nature of such differences and realistic sample sizes; the main evidence for this claim emerges from a series of straightforward simulations. Given this, the failure to find statistical signals of such hypothesized environmental differences cannot provide evidence against the existence and importance of such environmental differences. There remains, therefore, no way to generate robust evidence in favor of

the hereditarian hypothesis, at least given the actual state (and plausible near-future states) of the world.

I do not wish to claim that either of these claims is particularly original. As will become clear, I take the first claim to be well-established, but argue that the importance of it to these debates has not been as fully appreciated as it might be, perhaps especially by those arguing in favor of the hereditarian hypothesis. Arguments in the literature have strongly suggested that the second claim is correct, but have not formally tested the sorts of models considered here. The simulations explored here therefore provide additional evidence that the statistical signals proposed are unlikely to be found, even if the proposed pathways supposed to generate those signals in fact operate.

II. Environmental Differences Within and Between Populations

To test the hypothesis that the Black/White "IQ gap" is due in part to differences in the average endowments of genes directly relevant to the development of the sorts of abilities measured by IQ and related tests requires that one be able to separate out the effects of the different environments experienced (on average) by members of the different populations from the effects of the different (average) genetic endowments of members of the different populations[4]. No one, on either side of the debate, seriously

[4] There is no consensus on the best way to think about how different, on average, the genomes of Black Americans are from that the genomes of White Americans. Without knowing more about what question, precisely, one is asking, it is not at all clear that the question is well enough formed to have anything like a single answer (see Winther and

Kaplan 2012). On one view, we should regard "race" in humans as having at best a trivial genetic component. Humans are a very *non*-diverse species, genetically; we differ by only about 1 nucleotide in 1000, perhaps an order of magnitude less than many otherwise similar species. *Most* of what little variation exists can be found *within* any given local population one cares to identify, and the "continental" populations often identified as "races" account for only part of the remainder; Lewontin's original estimate, despite suffering from a number of technical problems and evoking a number of questionable assumptions, was in the end fairly accurate, and recent estimates are similar – about 85% of total genetic variation is within populations, with the remaining 15% split between variation between populations within the major continental "races," and variation between races identified with "continent of origin". But small differences are not no differences, and those small differences in e.g. allele frequencies associated with population structure can be exploited by contemporary population genetics (along with sequencing technologies and sophisticated computer programs like STRUCTURE) to sort individuals into "clusters" based on their likely relationships; given particular choices about the number of clusters to use, some of the populations discovered resemble some of the populations socially recognized, including what many would argue are recognizably the "races" identified in contemporary U.S. social discourse. Similarly, when enough alleles are tested, individuals can be assigned to particular clusters with great confidence. Whether this means that "races" as usually understood are biologically legitimate entities is, however, still subject to debate. For discussion, see Kaplan and Winther 2012 and cites therein, Winter and Kaplan 2013 and cites there, and Kaplan 2011).

doubts that the average environments experienced by Black Americans is markedly different (and, in almost every way relevant to well-being, worse) than the average environment experienced by White Americans; as Jensen himself realized in his original 1969 foray into this topic, it is this difference in the average environments experienced that prevents one from moving directly from the high heritability estimates for performance on IQ tests within some identified population to some conclusion regarding the likely differences in average realized performance between the groups being a direct result of some differences in average genetic endowments (Jensen 1969 pp 70, 81ff).

Flynn (for whom the "Flynn effect" is named) thought that the arguments over the hereditarian hypothesis came down, essentially, to an argument about "what would happen to the mean IQ of American Blacks if they found themselves distributed among the range of environments existent in contemporary White America in the same proportion as Whites" (Flynn 1980 73)[5]. Given the impossibility of actually performing a

[5] Of course, such a distribution would *not* immediately eliminate the effects of past racism, nor even do so in a single generation, given that some of those effects are biologically inherited via maternal imprinting / fetal programing, including for example heritable changes in the epigenome (see e.g. Novakovic and Saffery 2012; Collins et al 2011). As we continue to learn more about the intergeneration transfer of biological harms originally caused by social mechanisms, including racism, we will have to continually re-evaluate what would constitute a "fair" test of the hereditarian hypothesis.

rigorous test in which such a distribution is actualized[6], the "traditional" approach to attempting to answering this question is to try to find ways to figure out what the average influence of particular environmental differences is within each population of interest, and compare the environments of Black and White Americans with respect to that kind of variable[7]. Since Jensen's first major foray into this area, hereditarian

---

[6] I note below that the ubiquity of racism renders rigorously determining what should count as "the same" environment impossible; more generally, however, "mere" practical and ethical considerations make such an experiment an obvious impossibility.

[7] In the literature, differences in environments like these are referred to as "VE" differences (see Sesardic 2000, following Jensen). When an analysis of variance is undertaken on the variation in performance on IQ scores in a population, the traditional terms deployed are V_G (broad-sense heritability), V_E (the variance associated with environmental variation), V_{GxE} (the variance associated with gene-by-environment interactions), and V_e (the variance unaccounted for by the other factors, including both "individual" environmental effects, and errors and other forms of 'noise'). The variance associated with "shared" environmental variation – V_E – can, it is argued, be detected in both the population as a whole, as well as within each subpopulation of interest, although discovering precisely what kinds of variations "do the work" of producing the variance observed is difficult. Since V_E factors have the same effects in both populations, their influence on the between-population differences in outcomes can, in principle, be discovered by uncovering the differences in frequencies with which those types of environmental conditions are encountered in the populations of interest. It is worth keeping in mind that it is fiendishly difficult to perform accurate analyses of

researchers have argued that, with respect to those environmental variables known to influence IQ scores, the average developmental environments faced by Black Americans are not sufficiently worse than those encountered by White Americans to account for the differences in IQ scores observed (see Jensen 1969 81-88, 1998 455-456; Herrnstein and Murray 1994 299; Sesardic 2000 589-590, 2005 138-140; Ruston and Jensen 2005a; Godfredson 2005; etc.). So for example, children growing up in homes owned by their parents perform better on cognitive tests than those whose parents rent, *ceteris paribus* (Haurin et al 2001), and it is well-established that Black Americans have lower rates of home ownership than do White Americans (see e.g. Callis and Kresin 2013); however, given the small effect size, the difference in home ownership rates is unlikely, alone, to account for much of the observed difference in outcomes.

Hereditarian researchers argue that in order to explain the "IQ gap" in purely environmental terms, the "average environment" of Black Americans must be truly terrible, compared to the "average environment" of White Americans, for the development of the abilities associated with performance on IQ tests (and related

---

variance of this sort in humans (and other systems where systematic experimental manipulations are impossible or difficult), and that the results of such analyses are "local" – limited to the particular environments, genotypes, and distributions actually tested (see Kaplan 2000); how important these caveats are to the use of these approaches for understanding variation in IQ scores between populations and over time remains controversial, but many researchers at least remain very skeptical of the usefulness of these kinds of approaches.

measures) (for discussion of this argument, and the ways in which it might be misleading more generally, see Flynn 2009 36-37, 106-107ff). If, for example, one accepts a "gap" in performance on IQ and related tests of about one standard deviation, then, given a "moderate" value for the heritability of performance on IQ and related tests (say, 40%), the average environmental difference between the populations with respect to those variables associated with the development of IQ test performance must be around 1.3 standard deviations to fully address the outcome difference; if heritability is around 80%, the difference must be around 2 standard deviations (see Jensen 1998 455-456, Herrnstein and Murray 1994 299, Sesardic 2000 589-590, 2005 138-140)[8].

Could it really be the case, these hereditarians[9] wonder, that the average developmental environment, with respect to the development of IQ test performance, for Black Americans, is equivalent to an environment in the bottom 10% (if we assume a "middle" heritability value), or even the bottom 5% (for a higher value), of White

[8] Keep in mind that the many researchers less convinced that high estimates of heritability are warranted (Plomin, who few would accuse of habitually understating the heritability of traits, has in recent years consistently cites estimates in the .4-.6 zone; see Trzaskowski et al 2013 1051) and less convinced that "1 standard deviation" remains a reasonable estimate for the "gap" is performance on IQ and related tests (e.g. Nisbett, see above), will naturally find this line of argument less compelling.

[9] Sesardic has not, to the best of my knowledge, explicitly stated in print that he is a hereditarian of the sort discussed above. Nevertheless, I include him among them as the main thrust of his work has been to defend the hereditarian position against (what Sesardic views as very poor) opposing arguments.

Americans?[10] This, they claim, at least with respect to those environmental variables known to be associated with differences in performance on IQ and related tests within populations, is implausible (Jensen 1998 455-456, Herrnstein and Murray 1994 299, Sesardic 2000 589-590, 2005 138-140).[11]

But this line of reasoning depends on our being able to meaningful compare the environments of Black and White Americans, such that we could plausible say of two families, one Black and one White, that the children in each family experienced "the same" environment, at least with respect to those environmental variables that are

[10] This way of framing this issue is in fact deeply misleading. We can really only make sense of these claims if we presume, naively, that the environmental variables in question have their effects on cognitive development in a straightforwardly linear way (for discussion, see Kaplan 2000 70-74), and that changes in these environments "mean" the same thing, with respect to cognitive development, in the different populations in questions (see below). Neither assumption is warranted.

[11] It is perhaps worth noting that with respect to *some* environmental variables, Black Americans are, on average, straightforwardly this poorly off, worries about linearity, etc., notwithstanding. So for example the median Black American now have less than 6% as much *wealth* as the median White American, and with respect to wealth, at least, the vast majority of Black Americans households are, depending on the numbers one chooses, certainly not much better off than the bottom 10% of White American households and possibly rather worse off (see e.g. Kochhar, Fry and Taylor 2011). Of course, the relationship between (a lack of) wealth and IQ test performance is not straightforward (see Orr 2003 for discussion).

associated with the development of those skills relevant to performance on IQ and related tests. But how plausible is this assumption? Consider a trivial example – do a Black and White child in the U.S. experience "the same" environment if they watch the same TV shows, under similar socio-economic conditions? Insofar as the TV shows they watch perpetuate racist stereotypes, or report news in racially biased way, or for that matter, accurately and fairly report news from a racially biased world, arguably, the two children do *not* experience a similar environment. One experiences an environment in which watching TV makes relevant their race (and does so in a distinctly negative way); the other an environment in which watching TV is, with respect to their race, broadly neutral.[12]

Here is one place that Steele and Aronson's work on "stereotype threat" can be usefully deployed. One important fact revealed by Steele and Aronson's research (Steele and Aronson 1995, 2004; Steele 1997, 1998, etc. See e.g. Nguyen and Ryan 2008 for review) is that apparently *trivial* changes in the testing environment can have profound influences on test performance *for members of some groups, but not for others*. In the case of interest here, the interventions had profound effects on the average test scores of Black students but not on the average test scores of White students. Claiming that the test taken was a measure of intelligence depressed Black scores (but left White scores unchanged); having students record their race on the exams itself depressed the scores of Black students (but left the scores of White students unchanged) (Steele and Aronson 1995). Steele and Aronson conclude that

[12] The literature on the ways in which the media portrays Black Americans is vast, but see for example Dixon 2008 on race and crime in news reports for one example.

aspects of the test that evoke particular "stereotypes" (poor Black academic performance, or poor performance on IQ tests, for example) create a situation that is uniquely stressful for Black students, but not for White students. An environmental change that is *irrelevant* for White students is *significant* to Black students. In short, an environment in which one has to check a box indicating one's race *means something different* to a Black American student than it does to a White American student.[13]

Similar examples could be compounded *ad nauseam*: Given the known disparity in "pre-text" stops by the police (see e.g. Harris 1999, Jernigan 2000, Alexander 2012 133-135 and cites therein) getting pulled over for a trivial (or non-existent) violation of a traffic law *means something different* to a Black driver than to a White driver. Given the

[13] How much of the "IQ gap" can stereotype threat, as revealed in these classic studies, and confirmed repeatedly, account for? There is no way to answer that question, at least in part because there is no way to completely eliminate the effect (do we really imagine that Black students given a test that did not involve checking a box with their race had forgotten that they were Black?). Nevertheless, the claim that the entirely, or even some significant portion, of the "IQ gap" might be due to stereotype threat understood in this narrow way is unreasonable, and is certainly not supported by the available evidence. Stereotype threat, in this context, is *not* meant to be an environmental explanation for the entirety, or even some large part, of the "IQ gap," but rather provide an example of the kind of environmental differences between populations that are easily missed in these analyses, and yet can have profound effects on performance. The provide, in other words, a dramatic demonstration of the power of what might otherwise seem like implausibly trivial "X-factors."

known disparities in the ways that Black and White customers are treated in retail establishments (see e.g. Gabbidon 2003; Harris et al 2005 and cites therein; Schreer et al 2009 and cites therein, etc.), being treated rudely or accused of theft by a clerk *means something different* to a Black shopper than to a White shopper. Given the existence of racial steering and housing discrimination more generally (whether conscious or not) by landlords, real estate agents, and the like (see Yinger 1995, 1997; Carpusor and Loges 2006; Roscigno et al 2009; etc.) being ignored, brushed off, or re-directed to a more "appropriate" neighborhood, *means something different* to Blacks searching for housing than to Whites. Given the known racial bias in hiring decisions (see e.g. Bertrand & Mullainathan 2004; Pager, Western and Sugie 2009; Pager, Western, and Bonikowski 2009; etc.), having ones résumé ignored *means something different* to Blacks engaged in job searches than it does for Whites.

In none of these cases can one "control" for the environmental variable by finding places where members of the different population experience similar environments, because the *meaning* of the environment depends on the race of the individual experiencing the environment, and *not* on the "external" features of the environment[14]. It simply doesn't matter that both Black and White Americans encounter rude sales clerks, for example (or get pulled over for trivial or non-existent traffic offenses, or have their résumés ignored, or…) because "the same" rude behavior *means something different* in

[14] Lewontin, in his 1983 / 1985, argued that this kind of difference in the ways that organisms perceived and made use of their environments was a kind of "construction" (1985 99), and part of what made individuating environments without reference to the particular organisms involved impossible.

the two communities[15]. While in principle one could compare, say, Black and White Americans who had roughly equal numbers of broadly negative encounters with the police over some period of time, doing so would not control for the variable of interest, because the negative encounters have different meanings for the two groups. It doesn't even matter that there is very likely some particularly unlucky White driver who has been subjected to more pretext stops than some particularly lucky Black driver; in one case, the pretext stops *lack* racial meaning, and in the other, no matter how rare they are, they are *suffused* with racial meaning. They are not the same kind of event, and the frequency with which they are encountered by some particular person doesn't matter to that.

The populations in question are best thought of as not simply encountering particular kinds of environments with different frequencies, but with encountering different kinds of environmental conditions entirely. It is for this reason that Flynn, writing in 1980, was mistaken when he claimed that:

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

[15] Note well that it is at least in part because of the *histories* of these kinds of unequal treatments, and the cumulative effects – both on individuals and on populations – of these kinds of unequal treatments, that they have very different meanings.

them does not vary significantly within both black and white America.

(Flynn 1980: 60)

Even if particular “effects” vary within both populations, the *meanings* that those effects have, because in the part of the history of how and why those effects are distributed within the populations, is very different. No White American is harmed by “poverty” in the same way as Black Americans are harmed, because the *meaning* of poverty, as well as its causes and history, are different in the two populations. Certainly many White Americans are *poorer* than many Black Americans, and their poverty has many horrible effects on the children growing up in those poor homes, but those effects may well be very different because the history and meaning of the poverty is different in the two cases. Even if, as suggested below, some of the harms of poverty, say, are shared within the two populations, some of the harms – namely, those associated with racism and history of race in the U.S. – will be unique to particular populations.

III. Searching for Evidence of “X-Factors”

In their discussion of Nisbett (2009), Rushton and Jensen (2010) summarize some of the claims Nisbett and others have made regarding the profoundly different environments experienced by Black and White Americans, all of which would, intuitively, seem to disadvantage Blacks with respect to average performance on IQ tests and related measures. These include: “poor prenatal care and nutrition, relative infrequency of breast-feeding, deficiency of vitamins and minerals, lead poisoning, fetal alcohol poisoning, poorer health care, greater exposure to asthma-causing pollution, emotional

trauma, poor schools, poor neighborhoods along with the less desirable peers who come along with the territory, and much moving and consequent disruption of education," as well as different parenting styles (associated in part with the above stressors), stereotype threat, involuntary minority / caste status, and a host of other factors (Rushton and Jensen 2010 16). These do, on the face of it, seem to form a reasonable sub-set of the kinds of factors to suggest as environmental causes of the "IQ gap"[16].

What is surprising, however, is that Rushton and Jensen do not then actually *address* these possible causes, per se. They do not, for example, explore or even cite literature on the effects of e.g. home ownership measures on children's cognitive

[16] Note that there are two distinct elements to many of these factors; a child growing up in a home with a single under-employed mother and an absent father will suffer a variety of disadvantages that will impact cognitive development, whatever their race. But the underemployment of the mother, say, is very different in kind; in one case, it is wrapped up in practices that predictably result in racist outcomes (whatever the intent), and in the other, it is not (e.g. Pager, Western, and Bonikowski 2009) Similarly, the absence of the father emerges from a very different sociological background; in one case, a background suffused with the effects of a long history of racism and the associated limitations on available opportunities, and in the other, it is not (see e.g. Flynn 2011). Rowe et al 1994 and 1995 show, to a first approximation, that the disadvantages due to the lack of income in the two cases have a similar impact; for the reasons explained in below, the research does not show that there are no *additional* impacts.

development (see e.g. Haurin et al 2001), and then argue that the expected effects are too small to matter, or misleading, etc. Instead, they claim that *if* any causes like this in fact were responsible (in part) for the "IQ gap," that this would be revealed by a statistical analysis that searched for differences in the effects that other factors have on the average IQ scores of Blacks and Whites (Rushton and Jensen 2010 16-17).

It is this kind of analysis that Sesardic is referring to when he argues that if there are some environmental differences that affect only one population ("X-factors"), since those environmental influences will vary within the population in which they occur, the effect of those environmental influences should be detectable. He writes:

> Even if the X-factor does not exert its influence via VE factors, there is still a way to detect its presence. Assuming that it varies within the minority group (surely not all blacks are exposed to the same degree of discrimination!), the X-factor would increase the phenotypic variance in the affected group as well as the variance in any variable "touched" by the X-factor. (Sesardic 2005 141)

Since no increased variance is detected, Sesardic continues, the "X-factor" hypothesis is, at best, problematic.

This is the same reasoning than Rushton and Jensen deploy in their response to Nisbett. They note that at least some of the kinds of factors Nisbett is drawing attention to are "X-factors" and claim that:

> One way to test whether hypothesized X factors are operating to lower IQ scores for Blacks on standardized tests is to compare the similarity of the

> correlations between background variables (such as the home environment or the peer group) and outcome measures (such as scholastic achievement or delinquency rates). If Black-specific X factors are truly having an effect, some of these correlations for Blacks should be offset (up or down). (Rushton and Jensen 2010 16)

Since several studies (Rowe et al 1994 and Rowe et al 1995) found "no distortions in the correlations between the background variables and the outcome measures," Rushton and Jensen conclude that there is no evidence for "any minority-specific developmental factor" (Rushton and Jensen 2010 16-17) (they make a similar argument in their 2005a 249-253).

The argument, in a nut shell, is that *if* there is some "X-factor" that is depressing Black IQ scores (but not the scores of Whites), then, assuming that this factor *varies* within the Black population, some Blacks will be exposed to greater, and some to lesser, "X-factor" related score-depression. Assuming that this variation is random with respect to what the distribution of scores *would have been* in the absence of the "X-factor," the overall variance of Black scores should be *higher*, since the new range of scores will have been created by randomly suppressing the scores of Blacks, but not Whites, by a *varying amount* from what the scores would *have been in* the absence of the "X-factor." Alternatively, if some variable is usually associated with some portion of the variance in IQ test scores in a population, the presence of an "X-factor" that increases the overall variance should *reduce* the proportion of the variance associated with that factor (if it is unrelated to the X-factor) or *increase* the proportion of the

variance associated with that factor (if it is correlated with the X-factor). Either sort of change, if detected, would be evidence of the existence of such a factor.

To see how this is supposed to work, imagine two populations, each of which has a normal distribution of scores around a mean of 100 with a standard deviation of 15 (for the sake of the parallel to IQ test results). If we suppress the average score of one population by 15 points, by reducing the score of each individual member of that population by between 0 and 30 points (with a continuous uniform – a flat – distribution, such that the average reduction is 15 points, but no value is more likely than any other), it is clear that the *distribution* of IQ scores in that population will look quite different; the mean will have shifted down to 85 (by design) but the variance – the spread of the scores – will have increased (in this example, the standard deviation will trend towards around 17, up from the original 15)[17]. This difference would be easily detectable as statistically significant even with fairly modest sample sizes[18].

[17] Two methods are used in this paper to analyze the effect of various hypothetical influences on these patterns. In the first, used mostly for qualitative analysis, a normal distribution with a population size of 1000, a mean of 100, and a standard deviation of 15 was generated (using Wessa 2008); the populations so-generated were then manipulated as described in the text using Excel. Multiple "runs" were then averaged to generate estimates of the average changes in the standard deviations and variance; differences in the population variances were tested for statistical significance (at the .05 level) via the F-Test in Excel. In the second, R (with the "car" package installed) was used both to provide greater statistical power and additional confidence in the results,

The argument endorsed by Sesaric, Rushton and Jensen, and others, essentially follows this reasoning. If, for example, the IQ scores of Blacks are being suppressed by an “X-factor” variable that varies within the Black population, but is absent from the White population, the distribution of test scores by Blacks should be different (have a larger variance) than that of Whites. Alternatively, such a variable should produce changes in the association other variables (family income, say) have with IQ scores. Since no such differences have been found, Rushton, Jensen, and Sesardic reject the hypothesis that there is any such factor, or at least regard the hypothesis as having been rendered less plausible.

IV. Missing Evidence of X-Factors and the Role of Assumptions

---

as well as to provide better quantitative estimates of the power of the statistical tests to detect differences in the variance; here, Levene's test of homoscedasticity of variances was used to test for statistical significant results (again, .05 was chosen); samples were rerun for 10,000 loops in these cases. (The R scripts used are available online in “additional materials”.)

[18] R-based simulations, as described above, show that for a sample size of 1000 in each population, the Levene test would reveal a statistically significant difference in 99.5% of the trials. For a sample size of 500 in each population, the difference would be significant in over 90% of the cases. It is worth noting that for a sample size of 100 in each population, however, the difference would only be found to be statistically significant in about 30% of the trials.

There are several lines of argument that suggest that the failure to find the sorts of differences described above should have been expected. The most obvious emerges from reflecting on the assumptions used to explore the effect of the so-called "X-factor" on the statistical relationships in question. In the above example, it was noted that one had to assume that the "X-factor" varied within the population in question. The reasoning, recall, was that if the "X-factor" was something like 'the effects of racism,' then, since surely not every Black in America was exposed to the exact same level of racism (or to racism that had the same influence on their performance on IQ and related test measures), some Blacks would have their performance on IQ and similar measures reduced by relatively more, and some by relatively less, than the "average" of 15 points[19].

But in running the simulations that revealed that for reasonable sample sizes, a statistically significant difference would generally be detectable, it was assumed not just that that those radicalized environmental differences associated with growing up and living as a Black person in America that influence performance on IQ (and related) tests

[19] The "15 point" assumption here is unnecessarily high, but is used for illustrative purposes. First, see footnote 2, above, on arguments surrounding the actual value of the current "gap" in the U.S. context. Second, note that no hereditarian currently claims that *no* part of the current gap, whatever that gap may be, can be accounted for via the usual mechanisms of VE-style environmental differences (see e.g. Rushton and Jensen 2005a 279).

varied between members of the relevant population, but that the effects of these environments could be treated as a *single* effect, an effect that varied from nearly absent (no effect) through very significant (a 30-point decline in average IQ score), with every reduction in score being equally likely - a uniform (flat) distribution. While the claim that every Black American experiences an environment that results in an exactly equal disadvantage (say, is exposed to exactly the same amount and kind of racism) is implausible, so too is the assumption of a uniform distribution. A Black child growing up in America is surely *not* equally likely to experience *no* racism, moderate racism, *extreme* racism, etc. Nor does it seem at all reasonable to think that "racism" is a single *kind* of effect, such that the same "thing" ("racism") is experienced by e.g. a child of poor Black Americans growing up in the rural south and a child of middle-class Black Americans growing up in a city in the Northwest.

If we reject a uniform (flat) distribution, there are two obvious ways in which one might simulate the influence of these racialized environmental harms to make the assumptions somewhat more realistic.

First, one could consider the possible effects on IQ score of multiple different "X-factor"-type environmental stressors, and the extent to which these would be expected to vary quasi-independently. No White Americans, I have argued, grow up in environments that are relevantly similar overall to *any* of the environments in which *any* Black Americans grow up; however, there are obvious and important differences in the environments in which different Black Americans grow up. Differences in parental SES, education, etc., will obviously be important to the kinds of racialized environments (including racism) encountered, as will the different experiences of growing up in the

broadly Southern states versus, say, the North-East, as will broadly urban, suburban, and rural divides, etc. The racism (and racialized environments more generally) faced by e.g. young Black men in poor urban centers is different from, and will be experienced differently then, the racism (and racialized environments more generally) faced by a young Black women attending an elite University. In each case, some kinds of environmental stressors will be more common (and/or have larger effects), and others less common (and/or have smaller effects). But neither environment, nor any of the racialized stressors, is relevantly similar to the sorts of environments faced by (any) White Americans.

So for example, surely not every Black American who is given an IQ test (or related standardized test) suffers the same amount of stereotype threat (see e.g. Nguyen and Ryan 2008), and there is some research that suggests that the extent to which Black American test scores are reduced by stereotype threat varies between particular demographics (see e.g. McKay et al 2003). But the risk of suffering stereotype threat, and the severity of the effect, are likely broadly independent of e.g. the effects of preterm birth associated with the severity of the *mother's* experiences of racism (see Collins et al 2011). Neither influence is present in White Americans, both vary within the Black community, and the effect of these factors likely vary broadly independently from each other.

We might simulate the effects of such multiple quasi-independent factors by, instead of subtracting some value for *one* environmental variable from members of one of the populations(where the value is chosen at random from a continuous uniform distribution of between 0 and 30 points), subtracting values for *multiple* variables, each

of which we treat as a uniform distribution[20], chosen so that the average value subtracted comes out to 15 points. The average reduction in the outcome will be 15 points for the population so-treated (again, by design), but the distribution of reduction will no longer be uniform overall (indeed, as the number of variables with small independent effects increases, it will come to approximate a normal distribution; see below). We might, for example, choose 3 variables, each of which varies between 0 and 10 points, or 6 variables, each of which varies between 0 and 5 points, or 10 variables, each of which varies between 0 and 3 points; in each case, given the uniform distribution of each variable, the average reduction in the score will be 15 points. Given these assumptions, as the table below shows, even with very large sample sizes one would be unable to reliably detect a difference in the variances (see table 1).

# of environmental variables	Number of Individuals per Population		
	500	1000	5000
3 (x 10 points)	7%	10%	30%
6 (x 5 points)	5%	5%	7%

[20] It is worth stressing that the assumption of a uniform (flat) distribution remains unrealistic, even in this modified case, but it chosen for illustrative purposes. Also, it is of course implausible that each factor would have the same size effect; more likely is that there multiple factors with, say, variable medium-sized effects and more factors with smaller effects, some of which can vary broadly independently from each other and others of which are more strongly correlated. But as long as there are sufficient factors that vary broadly independently, the basic logic of the simulation holds.

10 (x 3 points)	4.5%	4.8%	5.5%

Table 1: Given multiple independent environmental X-factors, each with a continuous uniform distribution, how often will a statistically significant (Levene's test, .05) difference in variance between the populations be detected?[21]

Alternatively, one might imagine that we can treat *all* the different environmental harms as inputs into one "X-factor," and that given the large number of different quasi-independent racial harms, that the overall effect could be treated as a single variable with a normal, rather than as a continuous uniform distribution, with a mean of 15 and some standard deviation to be determined. If we imagined the variance within this combined X-factor to be the result of many many factors with small independent effects, we might set the standard deviation very low, say to 1. If we imagined that the variance would be similar in magnitude to the variance of realized performance on IQ tests in general, we would set the standard deviation fairly low (say, at around 2.25 or so). If we imaged instead that much of the factor was the result of relatively few, highly correlated inputs with large effects we would set the standard deviation fairly high. In any case, we can explore how changing assumptions about the distribution and variance. Again, for most reasonable assumptions and realistic sample sizes, differences in the variance will not be reliably detected (see table 2).

[21] Simulations were run on R, as described above. The percentages given are how often the Levene test revealed a difference significant at the .05 level between the variance of the population treated as described, and an untreated population. Keep in mind that in setting .05 as the cut-off, results close to 5% are indistinguishable from the expected false positive rate.

Standard Deviation of Distribution	Number of Individuals per Population		
	500	1000	5000
1	4.8%	5.2%	5.4%
2.25	5.9%	6.0%	12%
3.5	8.5%	13%	42%
5	20%	34%	93%

Table 2: For a combined X-factor that varies in the population in a normally distributed way, how often will a statistically significant difference in variance between the populations be detected?

That the correlation matrices presented by Rowe (1994, 1995) do not vary in statistically significant ways between the populations tested is supposed, according to Rushton and Jensen, provide evidence that there are no environmental stressors unique to Black Americans that could be responsible for any significant part of the so-called IQ gap (Rushton and Jensen 2010 16-17). But given the particular populations studied, and the sample sizes available, the particular statistical methodologies deployed have somewhat limited power to detect even known extant effects, and so it is known that even effects of fairly modest sizes would likely be missed.[22] As the above

[22] Dickens and Flynn 2002 make a similar point regarding the low power of actual studies to detect changes of this sort. It is worth noting as well that Hanscombe et al's (2012) analysis of the sample size necessary to detect fairly substantial changes in heritability, and the associated changes in variance, associated with GxE interactions,

shows, a variety of hypothetical pathways invoking "X-factors," none of which are terribly implausible, can account for both the entire IQ gap, and the failure of searches associated with changes in variance to detect statistically significant differences[23].

This result should not come as a surprise. It is worth reflecting on the those places where radical differences in IQ test performance are known to not reflect genetic differences, but where no particular environmental difference sufficient to explain to change is readily identified. There is, for example, no evidence of an increase in overall variance, nor in the association with other variables, associated with the increase in IQ scores within particularly populations over time, nor is at all clear what, precisely, the particular environmental changes that wrought those increases actually were (the

---

put the minimum sample size at about 5000 twin-pairs, over an order of magnitude larger than the sample sizes deployed by Rowe (1995) and substantially larger than those deployed by Rowe (1994), although the differences in the methodologies deployed limit the value of such comparisons.

[23] The statistical technique deployed by Rowe – comparing correlated vectors – looks for differences in the covariances of a range of variables. Rowe et al stress that an advantage of this approach is that it does not require specifying a model; however, testing the sensitivity of the approach *would* require specifying a model, and the sensitivity of this approach, compared to the sensitivity of simply looking for differences in the overall variances, will depend on the details of how the (hypothesized) "X-factors" interact with the identified variables (generally supposed to be "VE-factors"). But there is no reason to suppose that it will in general be *more* sensitive.

"Flynn" effect) (see e.g. Dickens and Flynn 2002 for discussion).[24] That there has been no obvious change in the variance in a case where, if the argument was correct, such a change should clearly be found is obviously problematic. This may mean that the effect in question is present but unidentifiable by the statistical tests employed (for example, the effect might be distributed in a different way than the hypothesis suggests, as suggested above), or it may mean that there is no effect to be found (again, see Dickens and Flynn 2002 for discussion). It implies, in other words, that there is something wrong with the reasoning that led us to expect such an effect. Whatever that may be, our failure to find such a signal in these cases also implies that our failure to find such effects in other, relevantly similar circumstances, *cannot* count as evidence in favor of the hereditarian hypothesis, given the *absurdity* of that hypothesis in these cases.

For the purposes of this argument, note that it doesn't matter whether the changes in IQ scores over time associated with the so-called "Flynn effect" reflect "real" changes in intellectual ability or not. Jensen, Rushton, and other hereditarians often claim that they do not reflect gains in "g" – general intellectual ability – and don't hold well for highly g-loaded tests (see e.g. Rushton and Jensen 2005a 269); others dispute

[24] Nor are any such effects obvious in the changes in IQ scores in nations that have undergone great economic growth or social change (e.g. East Germany, Roivainen 2012), though here it would be more fair to say that no one has looked closely for changes. Similarly, there is no evidence that regional differences in IQ scores within the U.S are associated with differences in the variance of scores, or associations between the scores and other variables, but, again, little work has been done searching for them.

this claim (see e.g. Colom et al 2001 and cites therein) or regard it is as irrelevant (see Flynn 2010 and cites therein). However, for the search for changes in variance associated with population specific X-factors, it doesn't matter whether the gains reflect "real" gains in cognitive ability or not. The gains in scores, whether they reflect "real" cognitive gains or not, must have a cause, and that cause must be environmental. If environmentally mediated gains leave a detectable signal, that signal ought to be clear in these cases, whether or not the gains in question are gains in "g" or something else entirely.

One might wonder, given the basic structure of the arguments above, why a change in variance would be expected for environmental but not genetic differences. The hereditarian position, recall, is that there are average genetic differences between members of the different "races," and these genetic differences are responsible for at least a large part of the IQ gap, via relatively direct effects that these genes have on the development of those intellectual abilities associated with performance on IQ tests and related measures. But unless those genetic differences are either *uniform* within the populations in question, or produce, via another mechanisms, a similarly uniform depression of scores, the same argument regarding an increase in variance, or a change in correlations with other environmental factors, ought to hold.

Here, one might reason that one should not expect to observe such changes associated with genetic differences because such differences are hypothesized to involve differences in *many* genes, each of which vary independently, and each of which has only a small effect. Certainly, if the hypothesized genetic differences are differences in genes that are associated with differences in intellectual ability within both

populations, this result must hold; there are simply no individual genes with large enough effects to matter alone. Despite large-scale and intensive searchers, very few alleles associated with differences in IQ test performance (within "White" populations) have been identified. While the contemporary estimates for the heritability of IQ test performance cited by hereditarian researchers range in the .5-.8 zone[25], searches have uncovered genes associated with no more than a tiny fraction of the variance (see e.g. Plomin 2013, Plomin et al 2013), and some researchers claim that most of these are likely false-positives (see e.g. Chabris et al 2012). Given the power of current methods, it is believed that any locus responsible for at least .5% of the variance ought to have been discovered, as well as at least many of the loci responsible for effect sizes of at least .1%, or somewhat less, of the variance (see e.g. Davis et al 2010; Plomin et al 2013). This has led some to the conclusion that if the current heritability estimates are to be believed, there must be at least hundreds, and perhaps thousands, of genes responsible for the variation in realized IQ test performance (and similar measures) within any particular population, each with a very small effect. If it is the relative frequencies of alleles of these genes that are supposed to vary between the populations then the net result would be a difference in realized IQ test taking performance with a very small standard deviation, resulting in a change to the distribution that would not be expected to change the variance in a detectable way.

[25] Again, these estimates are controversial; as noted above, Plomin currently argues for numbers in .4-.6 range, and some researchers arguing that much lower estimates are reasonable (see e.g. Nisbett 2009 and cites therein).

In the end, the failure to find a change in the variance or a change in the association between variables speaks only against prima facie relatively unlikely versions of either the hereditarian or environmental hypotheses; this approach is fundamentally unable to distinguish between more reasonable versions of the two hypotheses under consideration. The hypothesis that is genetic differences that are doing the work needs to posit differences in many (perhaps hundreds or thousands) of genes, each with tiny effects. But positing, without any evidence, systematic differences in hundreds or thousands of genes of small effects is surely no more plausible than positing multiple environmental differences with small effects! Indeed, as noted above, we do know of *some* environmental differences between the populations that are verifiably associated with differences in performance on IQ tests and related measures; we know of *no* genes that are so-associated.

V. Testable and Untestable Hypotheses

It is impossible to systematically control for the different environments experienced by different populations in the U.S. One cannot find e.g. Black and White children who have in fact grown up under “the same” environmental conditions, because no such children exist. Nor can one discover the effects of various differences in conditions within each population, and extend those results to between-population differences, because there are between-population differences that are not reflected within each population. Must, then, hypotheses regarding possible environmental causes remain equally untestable? Not quite.

Active manipulation of environmental variables suspected of being causally related to differences in the development of the cognitive abilities associated with performance on IQ tests is, of course, eminently possible. Indeed, making those kinds of changes is *precisely* what those who regard the gaps in average performance between populations as a social justice issue hope to do[26]. But insofar as current racism, and the effects of past racism, are responsible for the gap between the average scores of Blacks and Whites (in the U.S., and more broadly), one must recognize that it is unlikely that the gap will be eliminable in the near-future. Racism is simply too ubiquitous, and the effects are too varied, systematic, and long-lasting, for there to be any real hope of eliminating its effects in the near-term[27]. But working to eliminate those effects, even knowing that failure, in the near term, is inevitable, remains the only decent course of action available to us.

---

[26] Lewontin's claim that "the only way we could answer" the question "How Much Can We Boost IQ

and Scholastic Achievement?" would be to try to boost IQ and scholastic achievement is apropos here (1993 35).

[27] Again, it is important to stress that given what is now being learned about maternal imprinting and epigenetic changes associated with stress, eliminating the effects of racism should be recognized as a multi-generational project; even if, magically, racism were to vanish tomorrow, social-economic resources were redistributed completely equally, and all newborns randomly distributed across the population, it would very likely *still* be at least several generations before the full effects of racism were eliminated! See note 5, above).

References:

Alexander, Michelle. 2012. *The New Jim Crow: Mass Incarceration in the Age of Colorblindness*. The New Press. New York.

Bertrand, Marianne and Sendhil Mullainathan. 2004. “Are Emily and Greg More Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination.” *The American Economic Review*. 94(4): 991-1013.

Callis, Robert R. and Melissa Kresin. 2013. “Residential Vacancies and Homeownership in the Third Quarter 2013.” U.S. Census Bureau News, Social, Economic, and Housing Statistics Division. CB13-173.

Carpusor, Adrian G., and William E. Loges. 2006. “Rental Discrimination and Ethnicity in Names,” *Journal of Applied Social Psychology* 36: 934–952.

Chabris, Christopher F., Benjamin M. Hebert, Daniel J. Benjamin, Jonathan Beauchamp, David Cesarini, et al. 2012. “Most Reported Genetic Associations With General Intelligence Are Probably False Positives.” *Psychological Science*. 23(11): 1314-1323.

Colom, Roberto, Manuel Juan-Espinosa, Luís F. García. 2001. “The secular increase in test scores is a ‘Jensen effect'.” *Personality and Individual Differences* 30: 553-559.

Davis, Oliver S. P., Lee M. Butcher, Sophia J. Docherty, Emma L. Meaburn, Charles J. C. Curtis, Michael A. Simpson, Leonard C. Schalkwyk, and Robert Plomin. 2010. “A Three-Stage Genome-Wide Association Study of General Cognitive Ability: Hunting the Small Effects.” *Behavior Genetics*. 40:759–767.

Dickens, William T. and James R. Flynn. 2002. “The IQ Paradox Is Still Resolved: Reply to Loehlin (2002) and Rowe and Rodgers (2002).” *Psychological Review.* 2002, 109(4): 764–771.

Dickens, William T. and James R. Flynn. 2006a. “Black Americans Reduce the Racial IQ Gap: Evidence From Standardization Samples.” *Psychological Science*. 17(10): 913-920.

Dickens, William T. and James R. Flynn. 2006b. “Common Ground and Differences.” *Psychological Science*. 17(10): 923-924

Dixon, Travis L. 2008. “Network News and Racial Beliefs: Exploring the Connection Between National Television News Exposure and Stereotypical Perceptions of African Americans.” *Journal of Communication*. 58: 321–337.

Flynn, James R. 1980. *Race, IQ and Jensen*. Routledge & Kegan Paul. London.

Flynn, James R. 2009. *What is Intelligence?* Cambridge University Press. New York.

Flynn, James R. 2010. “The spectacles through which I see the race and IQ debate.” *Intelligence* 38: 363–366

Flynn, James R. 2011. “Black Youth: The Lost Boys.” in *African American Children And Mental Health, Volume 1: Development and Context*, edited by Nancy E. Hill, Tammy L. Mann, and Hiram E. Fitzgerald. Praeger. Santa Barbara, CA. 29-62.

Gabbidon, Shaun L. 2003. “Racial Profiling by Store Clerks and Security Personnel in Retail Establishments: An Exploration of ‘Shopping While Black’.” *Journal of Contemporary Criminal Justice.* 19(3): 345-364.

Hanscombe, Ken B., Maciej Trzaskowski, Claire M. A. Haworth, Oliver S. P. Davis, Philip S. Dale, and Robert Plomin. 2012. “Socioeconomic Status (SES) and Children’s Intelligence (IQ): In a UK Representative Sample SES Moderates the Environmental, Not Genetic, Effect on IQ.” PLoS ONE 7(2): e30320. doi:10.1371/journal.pone.0030320

Harris, Anne-Marie G., Geraldine R. Henderson, and Jerome D. Williams. 2005. “Courting Customers: Assessing Consumer Racial Profiling and Other Marketplace Discrimination.” *Journal of Public Policy & Marketing* 24 (1): 163–171.

Harris, David A. 1999. “Driving While Black: Racial Profiling On Our Nation's Highways.” *An American Civil Liberties Union Special Report*. June 1999. http://www.aclu.org/racial-justice/driving-while-black-racial-profiling-our-nations-highways

Herrnstein, R. J. and C. Murray. 1994. *The Bell Curve: Intelligence and Class Structure in American Life.* Free Press. New York.

Haurin, Donald, Toby Parcel, and R. Jean Haurin. 2001. “The impact of home ownership on child outcomes.” In *Low-Income Homeownership: Examining the Unexamined Goal,* edited by Nicolas Paul Retsinas and Eric S. Belsky. The Brookings Institution Press. Washington, D.C. 427-446.

Jensen, Arthur R. 1969. “How Much Can We Boost IQ and Scholastic Achievement?” *Harvard Educational Review* 39(1): 1-123.

Jernigan, Adero S. 2000. “Driving While Black: Racial Profiling in America.” *Law and Psychology Review* 24: 127-138.

Kaplan, Jonathan. 2000. *The Limits and Lies of Human Genetic Research*. Routledge. New York.

Kaplan, Jonathan Michael and Rasmus Grønfeldt Winther. 2012. “Prisoners of Abstraction? The Theory and Measure of Genetic Variation, and the Very Concept of ‘Race’.” *Biological Theory*. Online First. July, 2012.

Kochhar, Rakesh, Richard Fry, and Paul Taylor. 2011. “Twenty-to-One: Wealth Gaps Rise to Record Highs Between Whites, Blacks, Hispanics.” Pew Research Center. http://www.pewsocialtrends.org/files/2011/07/SDT-Wealth-Report_7-26-11_FINAL.pdf accessed 12/18/2013.

Lewontin, Richard C. 1983. “The organism as the subject and object of evolution.” *Scientia*. 188: 65-82. Reprinted in Levins, R and RC Lewontin, *The Dialectical Biologist*. Harvard University Press. Cambridge, MA. 85-108.

Lewontin, Richard C. 1993. *The Doctrine of DNA: Biology as Ideology*. New York, NY: HarperPerennial

.

Lynn, Richard and Tafu Vanhanen. 2002. *IQ and the Wealth of Nations*. Westport, Connecticut: Praeger.

McKay, Patrick F., Dennis Doverspike, Doreen Bowen-Hilton, and Quintonia D. McKay. 2003. “The Effects of Demographic Variables and Stereotype Threat on Black/White Differences in Cognitive Ability Test Performance.” *Journal of Business and Psychology*. 18(1): 1-14.

Nisbett, Richard E. 2009. *Intelligence and How to Get it: Why Schools and Cultures Count*. W.W. Norton and Company. New York.

Nguyen, Hannah-Hanh D. and Ryan, Ann Marie. 2008. “Does Stereotype Threat Affect Test Performance of Minorities and Women? A Meta-Analysis of Experimental Evidence.” *Journal of Applied Psychology*. 93(6): 1314–1334.

Orr, Amy J. 2003. “Black-White Differences in Achievement: The Importance of Wealth.” *Sociology of Education*. 76 (October): 281–304.

Pager, Devah, Bruce Western, and Bart Bonikowski. 2009. “Discrimination in a Low-Wage Labor Market: A Field Experiment.” *American Sociological Review*. 74(October):777–799.

Pager, Devah, Bruce Western, and Naomi Sugie. 2009. “Sequencing Disadvantage: Barriers to Employment Facing Young Black and White Men with Criminal Records.” *Annals of the American Academy of Political and Social Sciences* 623(May):195-213.

Plomin, Robert. 2013. “Child Development and Molecular Genetics: 14 Years Later.” *Child Development*. 84(1): 104–120.

Plomin, Robert, Claire M. A. Haworth, Emma L. Meaburn, Thomas S. Price, and Oliver S. P. Davis. 2013. “Common DNA Markers Can Account for More Than Half of the Genetic Influence on Cognitive Abilities.” *Psychological Science*. 24(4): 562-568.

Roscigno, Vincent J., Diana L. Karafin, and Griff Tester. 2009. “The Complexities and Processes of Racial Housing Discrimination.” *Social Problems*. 56(1): 49–69.

Roivainen, Eka. 2012. “Economic, educational, and IQ gains in eastern Germany 1990-2006.” *Intelligence*. 40: 571-575.

Rowe, David C. 2005. “Under the Skin: On the Impartial Treatment of Genetic and Environmental Hypotheses of Racial Differences.” *American Psychologist*. 60(1): 60-70.

Rowe, David C., Alexander T. Vazsonyi, and Daniel J. Flannery. 2004. “No More Than Skin Deep: Ethnic and Racial Similarity in Developmental Process.” *Psychological Review*. 101(3): 396-413.

Rowe, David C., Alexander T. Vazsonyi, and Daniel J. Flannery. 2005. “Ethnic and Racial Similarity in Developmental Process: A Study of Academic Achievement.” *Psychological Science* 1995 6(1): 33-38.

Rushton, J. Philippe and Arthur R. Jensen. 2010. “Race and IQ: A Theory-Based Review of the Research in Richard Nisbett’s *Intelligence and How to Get It*.” *The Open Psychology Journal* 3: 9-35.

Rushton, J. Philippe and Arthur R. Jensen. 2005a. “Thirty Years Of Research On Race Differences In Cognitive Ability.” *Psychology, Public Policy, and Law* 11(2): 235–294.

Rushton, J. Philippe and Arthur R. Jensen. 2005b. “Wanted: More Race Realism, Less Moralistic Fallacy.” *Psychology, Public Policy, and Law* 11(2): 328-336.

Rushton, J. Philippe and Arthur R. Jensen. 2006. “The Totality of Available Evidence Shows the Race IQ Gap Still Remains.” *Psychological Science*. 17(10): 921-922.

Schreer, George E., Saundra Smith, Kirsten Thomas. 2009. “‘Shopping While Black’: Examining Racial Discrimination in a Retail Setting.” *Journal of Applied Social Psychology* 39(6): 1432–1444.

Sesardic, Neven. 2000. “Philosophy of Science That Ignores Science: Race, IQ and Heritability.” *Philosophy of Science* 67(4): 580-602.

Sesardic, Neven. 2005. *Making Sense of Heritability* Cambridge University Press. New York.

Sesardic, Neven. 2010. “Race: a social destruction of a biological concept.” *Biology and Philosophy*. 25:143-162.

Steele, Claude M. 1997. “A Threat in the Air: How Stereotypes Shape Intellectual Identity and Performance.” *American Psychologist*. 52(6): 613-629.

Steele, Claude M. 1998. “Stereotyping and its Threat Are Real*.” American Psychologist* 53: 680-681.

Steele, Claude M. and Aronson, Joshua. 1995. “Stereotype Threat and the Intellectual Test Performance of African Americans.” *Journal of Personality and Social Psychology*. 69(5):797-811.

Steele, Claude M. and Aronson, Joshua. 2004. “Stereotype Threat Does Not Live by Steele and Aronson (1995) Alone.” *American Psychologist* 59(1): 47–55.

Trzaskowski, Maciej, Philip S. Dale, and Robert Plomin. 2013. “No Genetic Influence for Childhood Behavior Problems From DNA Analysis.” *Journal of the American Academy of Child & Adolescent Psychiatr.* 52(10): 1048-1056.

Wessa P., (2008), Random Number Generator for the Normal Distribution (v1.0.8) in Free Statistics Software (v1.1.23-r7), Office for Research Development and Education, URL http://www.wessa.net/rwasp_rngnorm.wasp/

Yinger, John. 1995. *Closed Doors, Opportunities Lost: The Continuing Costs of Housing Discrimination* (New York: Russell Sage Foundation).

Yinger, John. 1997. “Cash in Your Face: The Cost of Racial and Ethnic Discrimination in Housing,” *Journal of Urban Economics.* 42: 339–365.