

Commentary

Submitted to *Open Differential Psychology* August 10, 2014

Published in *Open Differential Psychology* August 25, 2014

The Elusive X-Factor: A Critique of J. M. Kaplan's Model of Race and IQ

Marc Dalliard¹

Abstract Jonathan Michael Kaplan recently published a challenge to the hereditarian account of the IQ gap between whites and blacks in the United States (Kaplan, 2014). He argues that racism and “racialized environments” constitute race-specific “X-factors” that could plausibly cause the gap, using simulations to support this contention. I show that Kaplan’s model suffers from vagueness and implausibilities that render it an unpromising approach to explaining the gap, while his simulations are misspecified and provide no support for his model. I describe the proper methodology for testing for X-factors, and conclude that Kaplan’s X-factors would almost certainly already have been discovered if they did in fact exist. I also argue that the hereditarian position is well-supported, and, importantly, is amenable to a definitive empirical test.

Keywords: IQ, *g* factor, race, racism, measurement invariance

¹ E-mail: mr_dalliard@hotmail.com

Contents

1. Introduction	3
2. Background	3
3. Kaplan's model	7
3.1. How racist is America?	8
3.1.1. Crime facts versus crime fiction.....	8
3.1.2. Race and policing.....	9
3.1.3. Labor market discrimination	10
3.1.4. Housing discrimination	12
3.2. Disparate treatment or disparate individuals?	13
3.3. The missing mechanisms	15
4. The nature of the IQ gap	16
5. Kaplan's simulations	18
5.1. Unimportance of variance differences	19
5.2. Abilities and tests	20
5.2.1. Measurement invariance	21
5.2.2. Stereotype threat.....	25
5.2.3. Flynn effect	26
5.2.4. Rowe and colleagues' findings.....	27
5.3. Can X-factors influence <i>g</i>?	28
7. Non-cognitive differences between blacks and whites	29
8. How to test HM	33
9. Discussion	34
References	37

1. Introduction

The hereditarian model (henceforth, *HM*) of the IQ gap between whites and blacks in the United States holds that the gap is mainly caused by genetic differences between the two races (Jensen, 1998; Rushton & Jensen, 2005). Kaplan (2014) challenges this view, arguing that racism and “racialized environments” are “X-factors” that can explain the gap. He presents simulations in support of this argument. He also claims that given the present state of science, there is no conceivable way to test HM.

I will show that Kaplan’s attack on HM is not convincing. My argument consists of four parts. First, I will show that Kaplan’s suggested explanation of the black-white gap is theoretically too vague and underdeveloped to be regarded as a serious model. Second, I will show that even if Kaplan’s model were to be considered as plausible, his simulations do not provide any support for it, or any evidence against HM. This is due to the fact that Kaplan ignores basic psychometric principles and most of the facts pertinent to any explanation of the gap. I will describe the proper method that could be used to search for X-factors. Third, I will argue that Kaplan’s model predicts that there are large racial differences in various non-cognitive traits, whereas such differences do not in fact exist. Lastly, I will show that, contra Kaplan, HM is a fully testable scientific model.

Before describing and dissecting Kaplan’s arguments, I will discuss the theoretical and conceptual background of the dispute.

2. Background

Arthur Jensen (1998, pp. 447–458; see also Sesardic, 2005, pp. 138–142) noted that there are two different models of environmental causation that could, in principle, explain the observed white-black IQ gap of about one standard deviation (15 IQ points):

- 1) According to the “variable environments” or *VE model*, all environmental factors influencing IQ are common to the black and white populations, but vary so that some factors are more frequent and others less frequent in one race versus the other. There are thus no factors unique to either race, but the black IQ disadvantage is caused by their having been exposed to more negative factors and/or fewer positive ones. The black distribution of environmental effects is shifted into the negative direction, with the average black growing up in a “cognitive environment” similar to that experienced only by disadvantaged whites.

- 2) The *X-factor model* is based on the idea that there are race-specific environmental factors that affect only one race. This is typically conceptualized as there being cognitively detrimental factors that affect all blacks and no whites. Thus the black IQ mean is lower than the white one because American society singles out all blacks for very specific IQ-sapping experiences. Jensen gave the name X-factor to the unknown non-genetic variable (or set of variables) that would affect the IQs of blacks but not whites.

If the VE model were true, it would mean that the environmental circumstances of the average black must be similar to those of the most deprived few percent of whites. The logic behind this calculation is the following. IQ has a high heritability within populations, perhaps as much as 80 percent in adults. If we assume that the genetic component does not cause racial differences, then the black-white gap must be entirely due to the environmental component, which accounts for as little as 20 percent of IQ variation. Environmental influences on IQ can be thought of as a unidimensional scale along which black and white individuals are distributed. Given that the total environmental effect on a given individual's IQ can be conceptualized as the sum of a number of more or less independent negative and positive factors, the distributions of the total environmental effects must be roughly normal. If the environmental influence on IQ variation is only 20 percent, then, for the VE model to hold, the mean of the distribution of environmental effects for blacks would have to be about 2.2 standard deviations lower than the mean for whites on the same scale of total environmental effects.² This would entail that the average black is exposed to a worse cognitive environment than about 99 percent of whites. Even if we assume that heritability is lower, say, 50 percent, the cognitive environment of the average black must be worse than that of about 92 percent of whites.

However, when black-white differences in the environmental factors that have traditionally been thought of as causes of the IQ gap have been investigated, it has been found that the differences are much too small to explain the gap. For example, differences in parental socioeconomic status can account for about one third of the gap (Herrnstein & Murray, 1994, p. 286), while according to Card & Rothstein (2007) residential segregation

² This is because the correlation between IQ and total environmental effects is equal to $\sqrt{0.20} \approx 0.45$, indicating that an environmental change of $1/0.45 \approx 2.2$ standard deviations is required to effect a phenotypic IQ change of one standard deviation, or 15 IQ points. See Jensen (1998, pp. 447–456) for a detailed discussion of this argument.

can explain about 25 percent of the SAT score gap (which is similar in size to the IQ gap). Similarly, Currie (2005) estimated that racial differences in health conditions explain at most 25 percent of the IQ gap in children.³ Phillips et al. (1998) found that even after controlling for more than 30 variables related to the economic, educational, cognitive, emotional, and health characteristics of parents and grandparents, about a third of the verbal IQ gap in children remained unexplained. The reason why it is very difficult to account for the gap in terms of environmental differences is that, firstly, they are usually not that strongly associated with IQ, and that, secondly, white and black distributions on those environmental variables overlap much more than what is expected on the basis of the VE model.

However, the problems of the VE model go much deeper. Eric Turkheimer has crystallized the results of many decades of behavioral genetic research into three “laws” (Turkheimer, 2000). They represent empirical generalizations of the causes of human behavioral differences. The first law states that all behavioral traits are heritable, while according to the second law familial resemblance in behavioral traits is mainly due to shared genetic rather than shared environmental influences. The third law states that the non-shared or within-family environment is an important source of behavioral differences. These laws apply to IQ, too, particularly after childhood as the heritability of IQ increases and shared environmental influences subside.⁴

It is important to understand that it follows from Turkheimer’s laws that proposed environmental effects on IQ are also expected to be confounded by genetic influences. Accordingly, behavioral genetic research indicates that “environmental” factors, such as measures of family environment, child rearing style, and peer relations, are under substantial genetic control (Plomin et al., 1994; Rowe et al., 1998; Kendler & Baker, 2007; Vinkhuyzen et al., 2010). Environments are not randomly distributed across the population, and an individual’s likelihood of encountering a specific environment may depend, in part, on his or her genotype or that of his or her parents, giving rise to spurious relationships between environmental factors and individual traits. “Partialing out” the influence of an “environmental” factor therefore typically also removes some of the genetic

³ This estimate is based on the unrealistic assumption that health problems are distributed independently of one another. Accounting for correlations between health problems would attenuate the explained variance.

⁴ Kaplan claims that the leading behavioral geneticist Robert Plomin has argued in his recent publications, such as Trzaskowski et al. (2013a), that the heritability of IQ is only in the range of 40–60 percent. In fact, Plomin has written that heritability in adults may be 80 percent, but that it is lower in children and adolescents (Plomin & Spinath, 2004). Trzaskowski et al. (2013a) analyzed a sample of 12-year-olds.

differences between individuals. Notably, the association between children's IQ and parental socioeconomic status appears to be mostly due to the influence of the same genes on both variables (Trzaskowski et al., 2014). Therefore, the reported correlations between IQ and environmental factors purporting to explain (some of) the IQ gap are, at best, overestimates of the true causal effects. Moreover, it is clear that the shared family environment is not a major cause of IQ differences within races, whereas the proposed environmental causes of the gap are generally shared between family members. Environmental influences on IQ are overwhelmingly non-shared (i.e., non-familial) in character, and few of them have been identified (Turkheimer & Waldron, 2000). Much of the "missing" non-shared influence on IQ may be *developmental noise* (Kan et al., 2010) that affects both races more or less equally, and therefore cannot contribute to the gap.

Simply put, it is not possible to explain the black-white IQ gap in terms of specific environmental differences because behavioral genetic studies indicate that very little of the IQ variation *within* races can be attributed to any identifiable environmental causes. While blacks grow up in environments that are in many respects inferior to those of whites, on the average, the distributions of environmental effects on IQ cannot be shown to greatly differ between races, which means that there is no credible evidence in favor of the VE model.

If there were an IQ-sapping environmental factor that harms all blacks but leaves whites intact, it could potentially explain the IQ gap even if VE-type factors cannot do so. The problem is, as Jensen pointed out, that it is difficult to come up with plausible candidates for such an X-factor. Any environmental influence on black IQ that one might think of would affect many whites, too. Significantly, the black IQ disadvantage is found across all regions, social classes, and generations; wherever one looks, blacks appear to suffer from a similarly sized IQ deficiency compared to white peers. This means that the putative X-factor must have very little variance, inflicting an almost constant 15 IQ point deficit on virtually all blacks. One might think that *racism against blacks* would be a perfect candidate for an X-factor, but a moment's reflection suggests that that racism much more closely resembles a VE-factor. More than thirty years ago, James Flynn articulated the inadequacy of racism as an X-factor in this way:

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

of them does not vary significantly within both black and white America.”
(Flynn, 1980, p. 60)

The conceptual implausibility of the X-factor model and the empirical inadequacy of the VE model lend credence to the hereditarian explanation.

3. Kaplan’s model

Kaplan rejects the argument that racism is not a promising X-factor. He thinks that the effects of racism on black IQ are not exhausted by the fact that racism may cause poverty, low self-esteem, or other VE-type effects. Instead, he suggests that the everyday experiences of just about all blacks, while superficially similar to those of many whites, are in fact infused by racism and are therefore qualitatively different in ways that affect IQ. He gives the following specific examples of the potential influence of racism or “racialized environments” on IQ:

- 1) Black criminals are overrepresented in television news, which means that the experience of black and white children viewing the same tv programming is different.
- 2) Blacks are more likely than whites to be stopped by the police for questioning.
- 3) Blacks who go into retail establishments are more likely to be suspected of theft or treated rudely by clerks.
- 4) Blacks face discrimination in the labor market.
- 5) Landlords, real estate agents, and other “gatekeepers” discriminate against blacks to keep them out of certain neighborhoods.

Kaplan suggests that such racialized environmental X-factors are prevalent in America, and that there are large numbers of them. He agrees that it is not plausible that their effect would be exactly the same on all blacks, but argues that if it is assumed that there are a number of uncorrelated X-factors, each with at most moderate variability, they would be very difficult to detect in a statistical analysis of test scores. He suggests that different classes of blacks are affected by different X-factors, but that all are affected to the same degree, causing a similarly sized IQ deficit in them regardless of class background and other VE-type circumstances. According to this model, the

racism encountered by, for example, “young Black men in poor urban centers” is different in its outward character but not in its IQ-sapping effects from that encountered by “young Black women attending an elite university.” Kaplan’s model is depicted in Figure 1.

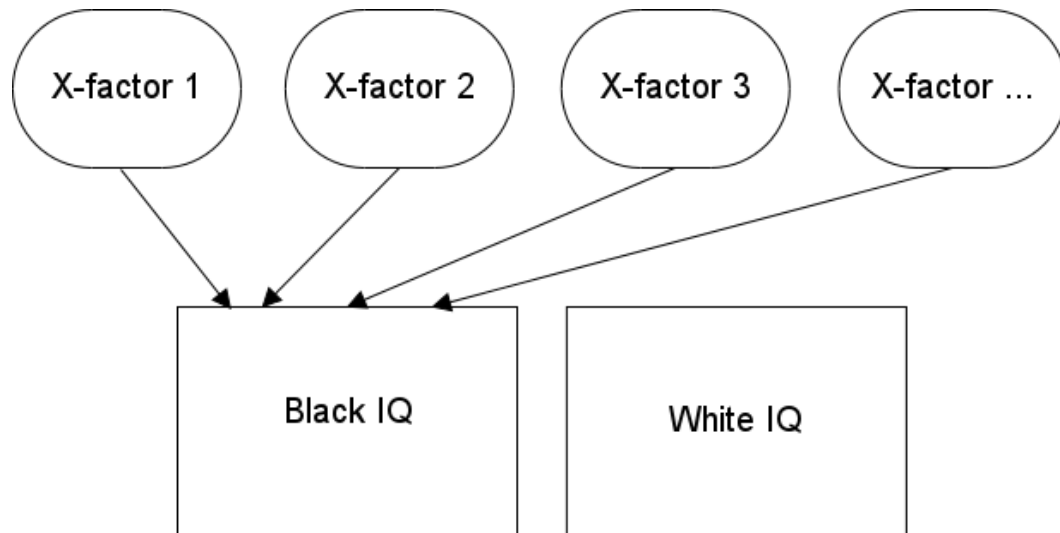


Figure 1. Kaplan’s X-factor model. Various environmental X-factors negatively influence observed black IQs, but have no influence on observed white IQs.

3.1. How racist is America?

While Kaplan appears to view the above list as clear-cut evidence of the pervasive influence of racism in American society, a closer look reveals that the evidence is ambiguous at best. The racial disparities discussed by Kaplan cannot be used as proof of racial bias or animus unless blacks are treated differently from non-blacks who behave the same way, and he offers no evidence that this is the case. I will next show how these examples of “racialized environments” can be plausibly interpreted in alternative, non-racial terms.

3.1.1. *Crime facts versus crime fiction*

Black criminals are generally overrepresented in television news coverage in relation to the black share of the population as a whole. However, there is little evidence of their being overrepresented in relation to the black share of the perpetrators of crime (Gilliam et al., 1996; Dixon & Linz, 2000; Chiricos & Eschholz, 2002). The fact that black offenders are disproportionately portrayed in crime news can be regarded simply as a

reflection of the great overrepresentation of black individuals in the ranks of criminals.

A better test of racial bias in television—and also more pertinent to Kaplan’s concerns about children’s television viewing—is the portrayal of race in fictional crime shows. Unlike news programs, such shows are not constrained by verisimilitude, which means that gross racial biases in the portrayal of crime are possible. However, studies have consistently found that compared to real-life crime statistics, blacks are *underrepresented* among criminal offenders in crime dramas, while whites are greatly *overrepresented* (Potter et al., 1995; Eschholz, 2002; Eschholz et al., 2004; Deutsch & Cavendar, 2008; Case, 2013). For example, 75 percent of the violent offenders and suspects in the 2000–01 season of *Law & Order* were white, whereas in the late 1990s only 13 percent of real-life violent crime suspects were white in New York City where the show was set. For black offenders and suspects, the proportions were 14 percent in the fictional world of *Law & Order* versus 51 percent in real life. (Eschholz et al., 2004, Table 1.)

3.1.2. Race and policing

Coviello & Persico (2013) found that while the New York City Police Department’s “stop-and-frisk” program led to blacks being stopped much more often than whites, the stops of whites were somewhat *less* “productive” in terms of arrests, which could be interpreted as evidence of a police bias against *whites*. To take another example, Worden et al. (2012) investigated vehicular stops made by the police in Syracuse, New York, over a period of four years, and found that African Americans were not more likely to be stopped during daylight than after dark when the police suffer an impaired ability to detect motorists’ race. This suggests that the greater propensity of black motorists to be stopped was not due to racial bias.

From these examples it is clear that racial disparities in encounters with the police do not constitute *prima facie* evidence of racial bias. Even if the police never relied on the (generally reasonably accurate) racial stereotypes about criminal offending, racial disparities in police scrutiny would arise because blacks are more likely than whites to engage in suspicious and illegal activities. The same inevitably applies to private security guards singling out seemingly disproportionate numbers of blacks for scrutiny. More generally, the observed black-white differences in crime rates are predictable from black-white differences in IQ and aggressiveness (Beaver et al., 2013), and victim surveys indicate that the high arrest and conviction rates of blacks reflect their genuinely high rates of offending (New Century

Foundation, 2005). The common belief that a racially biased criminal justice system underlies the high black crime rate is difficult to reconcile with these findings.

3.1.3. Labor market discrimination

Racial discrimination in the labor market is another area where Kaplan jumps to unsupported conclusions. He cites experimental audit studies where employers were found to prefer white job applicants to black ones with identical qualifications, arguing that this proves racial discrimination to be pervasive. However, Heckman (1998) has identified many severe limitations in this research. First, the experimental designs of such studies are based on dubious and untestable assumptions. Second, even if the experiments do identify genuine discrimination, the typical study reports only small differences between races, explaining very little of the existing racial disparities in the labor market. Third, the effect of racial discrimination on labor market outcomes is ultimately not determined by discriminatory employers but by those that actually employ blacks.

In a typical audit study, white and black “auditors” with matching (fictitious) credentials apply to low-skill, entry level positions, with the consequence that the studies have very poor ecological validity with respect to the labor market as a whole. The auditors sometimes exist only on paper, but experiments where actual persons are sent to job interviews are neither randomized (race cannot be assigned to individuals) nor double-blind (the auditors know the purpose of the study), which compromises any attempt to make causal inferences. The auditors can never be matched on all the variables that different employers may find important. It is often quite reasonable to regard white applicants as more qualified than ostensibly similar blacks. For example, the average IQ gap between black and white applicants to low-complexity jobs is 0.86 standard deviations, favoring whites (Roth et al., 2001), something that audit studies do not adjust for. Such racial differences may assume a greater-than-usual importance in the decision-making of the audited employers because many other characteristics that normally show racial differences in the applicant population have been experimentally equalized. In recruitment to cognitively more complex occupations, a rational employer would similarly expect a white graduate from a selective college to be smarter and more diligent than a black graduate from a similarly prestigious school, given the widespread use of racial preferences in college admissions.⁵

⁵ Espenshade & Radford (2009, p. 92) found that, after controlling for a host of other variables (e.g., high school grade point average and class rank, National Merit Scholar

A basic problem with many claims of group discrimination in modern, free labor markets is that they are based on the assumption that employers voluntarily leave money on the table. If the labor of some group were systematically undervalued by discriminatory employers, then surely some rational employers would step in and make a large profit on the basis of this market inefficiency. This would increase the demand on the labor of the discriminated-against group, driving up its wages. Widespread and significant labor market discrimination can continue only if there are legal or social norms that enforce discrimination even at a substantial economic cost to employers, but, as discussed below, such norms in today's America encourage or mandate discrimination in *favor* of blacks. Significantly, the measured job performance of black employees is inferior to that of whites working in similar occupations (Roth et al., 2003), whereas the discrimination thesis predicts the opposite. There is no evidence that the labor of black employees is undervalued in today's America.

Racial differences in labor market outcomes are clearly driven by "pre-market" factors, such as differences in education and IQ. When blacks and whites are equated on even a limited set of relevant pre-market factors, differences in their labor market outcomes are greatly attenuated or eliminated (Johnson & Neal, 1998; Carneiro et al., 2005). Indeed, the black-white income gap is often reversed after such equating (Johnson & Neal, 1998; Nyborg & Jensen, 2001; Heckman et al., 2006), which may signal the presence of discrimination in *favor* of blacks. While it is difficult to establish whether anti-black discrimination plays any significant role in the labor market outcomes of today's blacks, pro-black discrimination must play such a role, considering that it is something that is openly, legally, and widely practised in the name of "affirmative action", "diversity", and so on. For example, more than 60 percent of private sector workplaces in the US had affirmative action plans as of 2002 (Kalev et al., 2006), while federal and state agencies are bound by numerous rules concerning racial diversity in their hiring and contracting (e.g., Office of Federal Contract Compliance Programs, 2002). In fact, the disparate impact doctrine entails that many employers must in practice discriminate in favor of blacks so as to avoid legal repercussions (Wax, 2011). When one recognizes the fact that differences in skills and human capital are the primary reason for racial disparities in labor market outcomes, while also appreciating the prevalence of preferential treatment for blacks, it is apparent that Kaplan's case for

status, athlete and legacy status), the SAT scores of blacks admitted to selective private colleges were 310 points (out of 1600) lower than those of admitted whites, on the average, corresponding to a gap of about 1.5 standard deviations.

employment discrimination as a differentiating factor between whites and blacks is not credible.

3.1.4. Housing discrimination

Finally, Kaplan mentions housing discrimination by landlords, real estate agents, and others. Even this paradigmatic example of racial discrimination turns out to be ambiguous when examined more carefully. While the practices mentioned by Kaplan may contribute to residential segregation by race, it is not clear that racial animus drives the ostensibly discriminatory practices.

The reasons why one would want to control who gets to move into a neighborhood include such interlinked considerations as preserving property values, keeping crime levels down, and maintaining the quality of local public schools. In the presence of imperfect information, a rational actor interested in preserving a prosperous neighborhood would prefer whites to blacks as home buyers and tenants considering that the presence of blacks is statistically associated with many or all of the negative indicators for neighborhood value. Blacks may therefore end up being disproportionately turned down even when there is no racist intent. Considering that housing discrimination by race is illegal and thus a risky course of action, it is unclear if blacks are truly discriminated in the housing market when one compares them to objectively similar whites.

To establish that a pair of black and white individuals are really comparable in all their relevant characteristics, it is not sufficient to match them on just a few variables. A good illustration of this is the fact that black and white borrowers with the same credit scores and current incomes are not equally creditworthy in terms of the probability of loan default. Blacks consistently default more often than whites after adjusting for such factors as payment and credit history and income (Ferguson & Peters, 1995; Laderman & Reid, 2008; Anacker et al., 2012). When one appreciates the fact that the distributions of many important personal characteristics are different in the black and white populations, with the means of the black distributions located lower than the means of the white distributions, it is easy to understand why black individuals are not truly as creditworthy as ostensibly similar whites, on the average. For example, as a result of the different income distributions of blacks and whites, the expected future income of a white individual is higher than that of a black individual who has the same income in a particular year (Sanandaji, 2009). Over time, the characteristics of individuals tend to regress toward population averages which differ between races.

When deciding on who gets to rent or buy in a given neighborhood, it is not just the characteristics of a particular individual that may influence the decision. The way in which the *family members* of a prospective renter or buyer are perceived may also have an impact. Racial differences in the distributions of various psychological traits mean that the relatives of even highly accomplished black individuals tend to be inferior in many of their personal characteristics when compared to the relatives of seemingly similar whites. For example, the average levels of cognitive ability and academic achievement of upper-middle class black children do not resemble those of white children of the same social class but rather those of lower class whites (Herrnstein & Murray, 1994, p. 288; “Why Family Income”, 2008).

The economic, social, and physical decay of many urban areas in the wake of swelling black populations and white flight was a defining feature of American race relations in the 20th century. Seen against this historical background, it is difficult to argue that whites’ (and other non-blacks’) concerns about the character of their black neighbors are irrational.

3.2. Disparate treatment or disparate individuals?

The fact that blacks face adversities with disproportionate frequency is consistent with the racism explanation, but it is not the only possible explanation. Controlling for black-white differences in what can be plausibly interpreted as causally prior variables shows that many, if not all, of the outcome differences that Kaplan attributes to racism can be parsimoniously explained in non-racial terms. The same would in all likelihood apply to any further examples of racialized environments that he could come up with. After adjusting for relevant covariates, it is in fact not infrequently the case that whites rather than blacks appear to be targets of discrimination.

Interestingly, while Kaplan thinks that the available research justifies very expansive claims about the prevalence and effects of racism in contemporary America, he expresses great scepticism about the results of human behavioral genetics. He claims that given the non-feasibility of experimental manipulations, it is “fiendishly difficult” to make any accurate estimates of the influence of genes and environments in humans.⁶ I think

⁶ Kaplan also faults behavioral genetic analyses for being “local”, by which he means that the validity of the results is “limited to the particular environments, genotypes, and distributions actually tested.” This criticism is frequently advanced as if it were a special limitation of behavioral genetics, when it is in fact a common feature of all research on human subjects. For example, if it were discovered that a particular teaching innovation boosts the school achievement of children in California by 10 percent, this environmental

Kaplan greatly underestimates the power of behavioral genetic research designs. While behavioral genetics is not properly experimental, it relies on the convergence of results from different natural experiment paradigms (e.g., twin, adoption, and GCTA designs) increasingly applied to many large and representative population samples from around the world. This, together with the field's robust basis in both quantitative evolutionary theory and the results of non-human breeding studies, enables stronger causal inferences than are possible in almost any other area of social or behavioral science. In contrast, the research on racial discrimination cited by Kaplan relies on simple correlational analyses and single-blind quasi-experiments suffering from poor ecological validity and omitted variable bias. The causal interpretations Kaplan gives to these studies immediately crumble under even very simple robustness checks, as detailed in previous sections. Nevertheless, Kaplan thinks that not only is it *not* "fiendishly difficult" to make causal inferences about the influence of racism in the absence of experimental manipulations, it is positively easy: in his conception, a zero-order correlation between race and a negative outcome, or a quasi-experiment that is in all respects pitifully rudimentary compared to those routinely conducted in behavioral genetics enables one to draw far-ranging conclusions about the effects of racism in America. Kaplan's epistemological double standard cannot be explained away by the fact that he only briefly and cursorily reviews research on racism in America. The studies he cites appear to be quite representative in terms of study designs that are prevalent in this area of research (cf., Pager & Shepherd, 2008).

Kaplan's sweeping condemnation of American society as imbued by anti-black racism is premature. Of course, he is not alone in making this error—the conviction that racism has great explanatory power is widespread in certain sections of American society despite the distinct weaknesses of the evidence behind this conviction. The individual differences approach reflected in HM presents a necessary corrective to such beliefs about racism: psychological differences within and between races explain many outcome differences, indicating that accusations of racism against various institutions are often misplaced. Only after a thorough appraisal of the origins and significance of racial differences in socially valued traits can racism be allotted its proper role in understanding American society. Uncovering the etiology of the black-white IQ gap is particularly important, given IQ's pervasive importance in modern society (Gottfredson, 2002).

effect would not be any more or less generalizable to Texas schools than a finding that the heritability of IQ in California is 80 percent would be generalizable to Texas.

3.3. The missing mechanisms

The very existence of the “racialized environments” proposed by Kaplan is doubtful, but even if we accepted that potential X-factors of this type exist (or at least are perceived to exist by most blacks), Kaplan’s model would still be highly inadequate. This is because he does not offer any reason to believe that such factors would influence cognitive ability (of all traits).

Why would a security guard’s suspicious gaze at a store, a police officer’s gratuitous stop-and-frisk search, or a suspicion that a CV or an offer to purchase a house⁷ was overlooked because of racial bias cause an individual’s IQ to plummet? Kaplan does not present even a hypothesized mechanism of how this could happen, let alone any evidence that it actually happens. Notably, he claims that his X-factors are very heterogeneous, with each section of the black community being affected to the same degree by partly different sets of X-factors. Thus we are to believe that there are numerous different X-factors that depress IQ scores in the same way, yet we know nothing about the actual mechanisms behind any of them. We do not even know if these X-factors are supposed to decrease IQ in a permanent manner (say, each stop-and-frisk lowers an individual’s IQ by 0.5 points for the rest of his or her life), or only temporarily (a black murderer on tv makes the viewer somewhat less intelligent for the next two weeks).

Nor does Kaplan’s model suggest any explanation for the *familiarity* of the black deficit, that is, the fact that the IQs of black individuals are predictable from the IQs of their relatives with the same accuracy that obtains for white individuals (Jensen, 1998, pp. 447, 467–471). This familiarity suggests that X-factors would have to be tightly linked to family background, while from Kaplan’s description they appear to be much more randomly distributed. Another problem is that Kaplan’s model assumes experiences of racism to be ubiquitous, whereas only 1.7 percent of today’s black adults report that they are frequently treated poorly because of their race. A slightly higher proportion of whites, 2.3 percent, report often receiving such treatment.⁸

⁷ The fact that blacks may end up living in lower-quality neighborhoods as a result of housing discrimination must in itself be regarded as a potential VE-factor rather than a potential X-factor, given that the characteristics of the neighborhoods where whites and blacks live surely vary and overlap.

⁸ These results are from my analysis of Wave IV public-use data from the National Longitudinal Study of Adolescent Health: <http://www.cpc.unc.edu/projects/addhealth>. The sample is nationally representative and the participants were between 24 and 34 years old in 2008–2009 when the Wave IV interviews were conducted. The variables used were H4MH28 (“In your day-to-day life, how often do you feel you have been treated with less respect or courtesy than other people?”) and H4MH29 (“What do you think was the **main** reason for these experiences?”). The cross-sectional grand sample weights for Wave IV (GSWGT4_2) were used in the analysis.

Kaplan brings up stereotype threat (Steele & Aronson, 1995) as an example of a subtle environmental influence that can have a large effect on IQ scores, arguing that his X-factors could be similar in nature. However, stereotype threat is based on a causal theory of how anxiety about confirming a stereotype about intelligence hampers performance on intelligence tests. There is thus a direct and immediate link, supported by experimental evidence, between poor test performance and the proposed causal factor—something that certainly cannot be said of any of Kaplan’s far-fetched propositions. Furthermore, if Kaplan’s X-factors were real, their influence on IQ would have to be, for reasons discussed later in this article, far more subtle than that of stereotype threat.

Kaplan’s X-factor model must be seen as a casual speculation rather than a well thought-out challenge to HM. It suffers from so many implausibilities and lacunae that it cannot provide a credible explanation of the black-white IQ gap. Nevertheless, the fundamentally vague and impressionistic (if not downright fantastical) character of the model will be disregarded in the following sections because it is instructive to examine why the simulations that Kaplan presents in support of his claims in fact provide no such support.

4. The nature of the IQ gap

Before discussing Kaplan’s simulations, a consideration of certain facts about the black-white gap is in order. While Kaplan seems to conceive of HM exclusively in terms of VE-factors and X-factors, this particular argument is in fact just one piece in the body of evidence supporting HM. There are equally or more interesting arguments that Kaplan completely ignores.

One of the most important discoveries made by Arthur Jensen in his research on the black-white IQ gap was the finding that its magnitude is not invariant across different tests but tracks their g loadings, or correlations with the latent general factor of intelligence. He devised the method of correlated vectors (MCV) to assess the strength of this association. In MCV analyses, a vector of g loadings from a test battery is correlated with a vector of the values of some other variable, such as the black-white gap on different tests. The MCV tests if the other variable’s association with test scores is driven by g or by other sources of variance that are orthogonal to g . Psychometrically, these other sources represent non- g factor variances, test specificities, and measurement error, but except for measurement error (which can be partialled out) the MCV usually cannot specify the nature of the non- g variance. If the MCV correlation is large and positive, it indicates that the association between test scores and the other variable is primarily due to g . Conversely, a large negative MCV correlation indicates that the

association is driven by non-*g* sources of variance. If the MCV correlation is close to zero, the association between test scores and the other variable usually reflects some complex combination of influences that may involve both *g* and non-*g* components. In practice, MCV analyses are often subject to false positives and false negatives, and meta-analytic aggregation of MCV results is required for reliable inferences.

In a meta-analysis of 149 tests from 15 test batteries, Jensen found an average correlation of 0.63 between the magnitudes of black-white gaps and *g* loadings (Jensen, 1998, pp. 377–378).⁹ What this means is that the better a measure of the *g* factor a given cognitive test is, the greater the black-white gap on it usually is. The significance of this *Jensen effect*, as the positive MCV correlations between *g* loadings and other variables are called, is that such effects are otherwise only found for strongly genetically influenced biological variables.¹⁰ Specifically, the *g* loadings and heritability coefficients of tests have been found to be intercorrelated moderately to highly in many studies (te Nijenhuis et al., 2014b; Rushton & Jensen, 2010). Jensen effects have also been detected for correlations between test performance and inbreeding depression, heterosis, and head size (Jensen, 1998, p. 419), the last being a highly heritable characteristic (Smit et al., 2010) robustly associated with IQ (Rushton & Ankney, 2009). In contrast, strong “anti-Jensen effects”, or negative MCV correlations between *g* loadings and other variables, have been reported for the environmentality coefficients of cognitive ability tests, that is, the complements of heritability coefficients (Rushton & Jensen, 2010); for the effects of retesting or practice on test performance (te Nijenhuis et al., 2007); and for the test score gains induced by the Head Start compensatory education programs (te Nijenhuis et al., 2014a). Similarly, the observed increases in the cognitive test scores of many populations across much of the last 100 years (the Flynn effect) are correlated at -0.38 with the *g* loadings of the tests (te Nijenhuis and van der Flier, 2013).¹¹ Notably, Flynn et al. (2014) found in a meta-analysis that “biological-environmental” effects, such as iodine deficiency and traumatic brain injury, have a strong negative influence on cognitive test performance,

⁹ Given that this correlation is necessarily attenuated by certain statistical artifacts, such as sampling errors in the selection of subjects and tests, the true correlation must be substantially higher (Jensen 1998, pp. 380–383).

¹⁰ The discussion here concerns the domain of variables that are causally antecedent to test performance. Positive correlations have also been found between *g* loadings and outcomes influenced by cognitive ability, such as job performance (MacDaniel & Kepes, 2012), but such Jensen effects are not considered here.

¹¹ Considering that the causal structure underlying IQ tests does not stay constant across generations, as discussed below, the significance of the anti-Jensen effect on generational IQ gains is unclear.

but that this effect is unrelated to g loadings (MCV correlation ~ 0). If the black-white IQ gap reflected environmental rather than genetic disparities, it would constitute a very unusual Jensen effect.

Research on Jensen effects indicates that g is mainly a genetic phenomenon, and that variables that are positively associated with g are biological variables that share genetic influences with g . This is underscored by the finding that the kinds of environmental effects, such as brain injuries, that directly affect the neurobiological substrate of cognition do not cause g -linked cognitive changes. The principally genetic nature of g has also been supported in multivariate behavioral genetic analyses where genetic influences on different cognitive abilities have been found to be largely common rather than ability-specific (Plomin & Spinath, 2004; Trzaskowski et al., 2013b; see also Panizzon et al., 2014 where it was found that genetic correlations between different tests and abilities can be best explained in terms of a hierarchical g factor model).

Kaplan does not consider the g -saturated nature of black-white cognitive differences at all, despite this finding's centrality to the debate. What this means is that his proposed explanation of the gap cannot account for the pattern of cognitive differences that is actually observed. It also means that his simulations, discussed in more detail below, are misspecified and, for this reason alone, do not provide evidence for or against any realistic model of racial differences.

It should be noted that one cannot nullify the importance of Jensen effects by simply denying the reality of the g factor as a source of cognitive differences. Regardless of the nature of g , environmental variables are differentially associated with g loadings than genetically saturated variables, and the black-white gap resembles genetic variables in this respect. Any alternative, non- g theory of intelligence must be capable of explaining why we see these consistent patterns of correlations between g factor loadings and other variables.

5. Kaplan's simulations

Kaplan presents a series of simulations of the effects of his hypothesized racialized environments on the IQs of blacks. He claims that the simulations show that such effects would generally not be statistically detectable in any study with a realistic sample size. He concludes that racism against blacks is therefore a promising explanation of the IQ gap, and that HM is not viable.

Unfortunately, Kaplan's simulations are psychometrically so flawed that they cannot provide evidence in favor of his model or against HM. The flaws can be summarized in the following three points:

- 1) The test for the equality of variances which Kaplan uses to test for the presence of X-factors cannot be used for that purpose.
- 2) There are well-established ways to model intelligence differences and methods that can be used to search for X-factors in the framework of such models, but Kaplan ignores them.
- 3) The simulations disregard the empirically observed pattern of correlations between g loadings and black-white cognitive differences.

I will next discuss these three points in some detail.

5.1. Unimportance of variance differences

Kaplan uses Levene's test for the equality of variances to investigate whether his simulated X-factors inflate IQ variances to a statistically significant extent. He finds that given realistic sample sizes, the increases in variances are not generally statistically significant. He regards this as the main finding of his study, and concludes that X-factors are therefore not generally detectable. This conclusion is completely unwarranted.

Given the abundance of data on black-white IQ differences, one could easily conduct a powerful meta-analysis of variance differences. For example, a 2001 meta-analysis of racial differences in general cognitive ability (Roth et al., 2001) had sample sizes in the millions, enabling very accurate estimation of population parameters. If the variances of black IQ scores were slightly but consistently higher than those of whites, a meta-analysis would show it with a high degree of statistical reliability. As it happens, the variances of IQ scores in blacks are typically *smaller* than those of whites. Jensen (1998, p. 353) found that black standard deviations are usually in the range of 11–14 IQ points, with a mean of 12, compared to the white standard deviation of 15 points.¹² This indicates that the outputs of Kaplan's simulations do not even approximate actual IQ data. One of the peculiarities of his article is that he does not examine variance differences in any real-life data sets.

However, a more important reason why Kaplan's simulation results do not support his conclusions is that differences in IQ variances could be due

¹² However, this has not been universally found to be the case (e.g., Murray, 2007). It is conceivable that insufficient sampling of black individuals from the full range of ability or floor effects in tests have artificially lowered the variance of black IQ in many studies.

to other causes besides X-factors. Specifically, one group could be inherently more variable than another group on a given phenotype. For example, the pigmentation of hair and eyes varies in Europeans much more than in black Africans, reflecting the fact that the genetic mutations causing this phenotypic diversity in Europeans arose, or at least became selectively advantageous, long after the evolutionary divergence of African and non-African lineages. Given that there is no *a priori* reason to expect different populations to have exactly the same “natural” IQ variances, Levene’s test, which assesses deviations from a null difference, cannot provide any useful evidence for or against the existence of environmental X-factors. The proper way to test for X-factors is discussed next.

5.2. Abilities and tests

The predominant view among psychometricians and the one that is adopted in this article is that individual differences in intelligence can be conceptualized in terms of a factor hierarchy with a third-level general factor (g), second-level broad ability factors, and first-level test-specific variation (Deary, 2012). Higher-level sources of variation exert a causal influence on the lower levels of the hierarchy. Observed test scores, whether they be full- or subscale scores, subtest scores, or item scores, are regarded as reflections of the latent abilities that underlie performance on all cognitive tasks.¹³

The distinction between abilities and test scores is completely ignored by Kaplan. His simulated X-factors directly influence observed, full scale IQ scores (see Figure 1 above). However, full scale IQ scores are typically composites of scores on different tests. The fact that any causal influence on test performance is almost inevitably differentially associated with different tests and abilities offers rich possibilities for testing for group differences in causal processes. There are standard methods for doing such analyses. In contrast, Kaplan’s simulations are based on manipulating single test scores, are focused on uninformative variance differences, and are not grounded in any realistic model of intelligence. This means that they tell us nothing about how difficult or easy it is to detect X-factors.

¹³ There are alternatives to the reflective hierarchical factor models considered in this article. For example, g could be regarded as a non-causal formative factor (van der Maas et al., 2014). However, regardless of how one conceptualizes intelligence, a proper test for X-factors will have to consider covariance (and mean) structures rather than just variances like Kaplan does. Moreover, any alternative intelligence model must be able to account for well-established empirical findings, such as the Jensen effects discussed previously.

5.2.1. Measurement invariance

A proper test of Kaplan's model would involve the specification of a causal model for test score differences where X-factors would influence observed test scores in blacks alongside underlying abilities, whereas in whites only the underlying abilities (and unique variances¹⁴) would influence test performance. The plausibility of such a model could then be investigated through an analysis of *measurement invariance* in the framework of multiple-group confirmatory factor analysis. The analysis would examine whether simulated variance-covariance matrices and mean structures produced by the X-factor model could be statistically distinguished from those produced by the same model without X-factors. The X-factor-free white model and the black X-factor model are depicted in Figures 2a and 2b, respectively.¹⁵

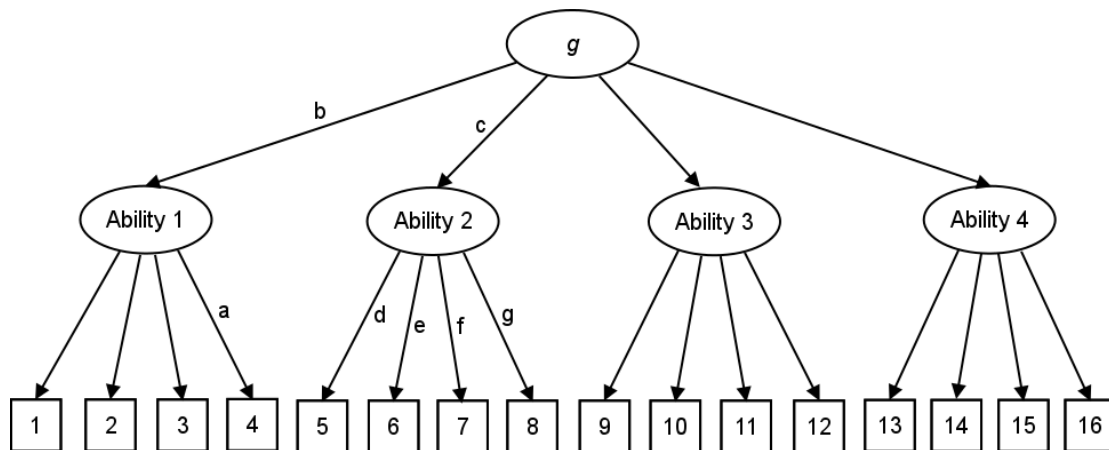


Figure 2a. Model for white test scores. The squares represent different cognitive tests, while the ellipses are latent ability constructs that, except for g , are unspecified here but could represent verbal, fluid, and spatial abilities, and short-term memory, for example. Residual variances are not shown but are assumed to be uncorrelated. The letters a – g are selected factor loadings.

¹⁴ Unique (residual) variances are influences on observed test scores that are independent of the ability factors. They comprise random measurement error and narrow abilities that are specific to a particular test. In hierarchical factor models, the lower-level factors have residuals, too, representing factor variance that is independent of the higher-level factor(s).

¹⁵ Kaplan alternates the number of X-factors and their variances in his simulations, but such details are unimportant here because his simulations lack evidential value and a proper test of Kaplan's model cannot presently be conducted, as explained in the main text.

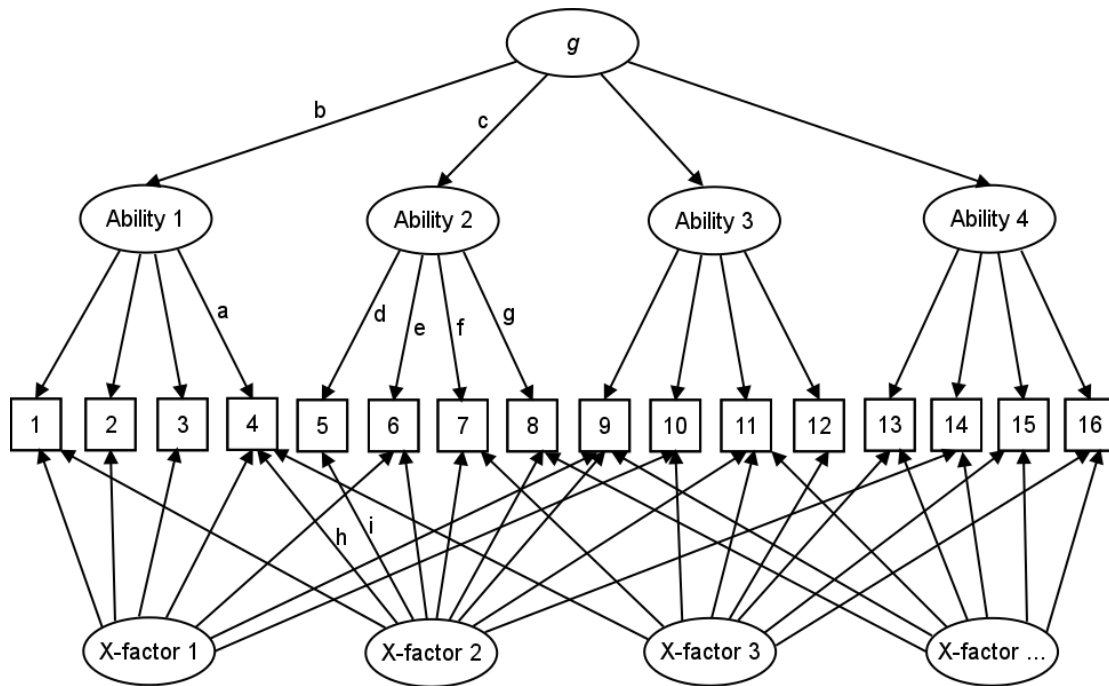


Figure 2b. X-factor model for black test scores. Various X-factors, conceptualized as latent variables, influence test scores alongside ability constructs. Residual variances are not shown but are assumed to be uncorrelated. The letters *a–i* are selected factor loadings.

The test for measurement invariance that could be performed on the simulated variances, covariances, and means produced by the white and black models would essentially be a test of whether it is statistically plausible that the black test scores that were actually produced by the X-factor model could as well have been produced by the white model. Assuming that the white model shows an adequate fit to the data generated by the X-factor model in a single-group confirmatory factor analysis (if it does not, the X-factors have already been detected and the analysis can end), we can proceed to a multiple-group analysis, in which the following four conditions are examined (Brown, 2006, pp. 269–270):

- 1) *Equal form.* Across the two groups, the number of latent factors must be the same, and the same tests must load on the same factors. This condition will necessarily be true if the white model fits the data produced by the X-factor model in a single-group analysis, but the equal form condition of the multiple-group analysis serves as a baseline model for the next step of the analysis.

- 2) *Equal factor loadings*. The loadings (or regression slopes¹⁶) of the tests on the factors must be equal across groups, that is, a change in the level of a factor must be associated with similarly-sized changes in the levels of the associated tests in both groups.
- 3) *Equal intercepts*. When the tests are regressed on their respective factors, the intercepts must be equal across groups. This guarantees that any differences in the means of the tests can be attributed to differences in the means of the factors. If the intercepts are unequal, it indicates that group differences in test means are not due to group differences in the underlying abilities.
- 4) *Equal residuals*. The magnitudes of the residual (unique) variances of the tests must be equal across groups. This ensures that any variance differences in the tests can be attributed to the latent factors.¹⁷

The plausibility of these four conditions is tested by sequentially introducing additional cross-group equality constraints on the model and examining whether the fit of the model deteriorates. If all the relevant parameters can be constrained to be equal across groups without a significant deterioration in model fit (compared to if the parameters were freely estimated for both groups), then *strict measurement invariance* holds across groups.¹⁸ Strict invariance indicates that test score differences between groups can be fully attributed to the same underlying abilities that cause differences within groups (Lubke et al., 2003).

It is easy to see how Kaplan's X-factors could violate measurement invariance. For example, in terms of factor loadings a_{-i} shown in Figures 2a and 2b, the model-implied correlation between tests #4 and #5, calculated

¹⁶ Factor analysis can be thought of as a type of linear regression analysis where observed test scores are regressed on latent factors.

¹⁷ Constraining residual variances to be equal is often considered to be an overly restrictive and unrealistic requirement, and this test is not always performed in analyses of measurement invariance (Brown, 2006, p. 290). However, residuals have previously been analyzed in the context of the white-black IQ gap (e.g., Lubke et al., 2003), and testing for their equality may be particularly informative when searching for X-factors. However, if it was found that the residual variances were the only non-invariant parameters in an analysis of X-factors, that should probably not be regarded as a disconfirmation of Kaplan's model.

¹⁸ Given that we are dealing with a hierarchical model, the invariance tests must be performed also on the higher-order part of the model so as to ascertain that the relations between the g factor and other factors are similar across groups. See Chen et al. (2005).

using path tracing rules, is $abcd$ in the white model and $abcd + hi$ in the black model.¹⁹ Because the measurement invariance model assumes that the black data can be explained using the parameters of the white model, the only way to account for the increased correlation (i.e., the term hi) between tests #4 and #5 in blacks is to make one or more of the factor loadings $a-d$ larger. This jeopardizes measurement invariance because equal loadings across groups is one of its requirements. Constraining the loadings to values between the optimal white and black ones may well lead only to a non-significant deterioration in model fit if just a few loadings are modestly affected. But as the X-factors introduce a large number of new dependencies between tests, many loadings will be affected, some strongly, making factor loading invariance difficult to achieve.

Another example of how X-factors can violate measurement invariance concerns the invariance model's assumption that group differences in the means of the tests can be explained by group differences in the means of the latent factors. This necessitates that test score gaps be collinear with factor loadings that are constrained to be equal across groups, that is, the size of group differences on different tests must be consistent with the size of the group-invariant factor loadings of those tests (Wicherts & Dolan, 2010). For example, the size of the black-white gaps on tests #5, #6, #7, and #8 in Figures 2a and 2b must be fully predictable from the size of the factor loadings $d, e, f,$ and g . This is tested by constraining the intercepts of the tests to be equal across groups, and examining whether the requirement to reproduce the mean differences in the tests from factor means leads to a deterioration in model fit compared to a model without this constraint. From Figure 2b it is apparent that multiple X-factors exert negative influences on the means of the tests in blacks in a way that is completely unrelated to the loadings of the tests on the ability factors. Therefore, X-factors tend to change the pattern of black-white gaps on different tests so that the gaps are no longer predictable from ability factor loadings, leading to non-invariant intercepts across groups.

If the tests for measurement invariance showed (across many iterations) that the black and white models of intelligence produce significantly different variance-covariance matrices and mean structures, this would indicate that the invariance tests successfully detect the existence of X-factors. If, on the other hand, there were no significant differences between the black and white matrices and mean structures, we would conclude that the method is not sensitive enough to detect X-factors. Unfortunately, Kaplan's model is psychometrically very underdeveloped, providing no

¹⁹ For simplicity's sake, the example uses standardized regression weights (correlations) even though an actual invariance analysis would use unstandardized parameter values.

information on how his X-factors would influence performance on different kinds of tests. Why is the black-white gap greatly attenuated on tests of short-term memory and perceptual speed, while it is particularly large on tests of general knowledge and abstract reasoning? Kaplan provides no explanation. Jensen explained such findings by reference to the varying *g*-loadings of cognitive tests, showing that controlling for the influence of *g* eliminates the vast majority of cognitive differences between the two races.²⁰ Given the lack of information on how Kaplan's X-factors would influence different tests, it is not currently possible to analyze if they could be detected using the procedure just described.

It has been repeatedly shown that black-white differences on IQ test batteries satisfy the requirements of measurement invariance (Dolan, 2000; Dolan & Hamaker, 2001; Lubke et al., 2003; Trundt, 2013). This indicates that the same latent abilities that explain test score differences within each race also explain the observed interracial IQ gap. The fact that statistical signals of race-specific X-factors are not empirically observed in the literature on IQ measurement invariance strongly suggests that Kaplan's model is a non-starter. He could, of course, aver that his X-factors are so subtle that they would not violate measurement invariance, but he has not tested this claim and it cannot presently be tested given the sketchy nature of his model.

5.2.2. *Stereotype threat*

However, we can get a good idea of whether a properly specified X-factor model would pass a test of measurement invariance by examining whether environmental factors known to influence test scores pass this test. In particular, Kaplan identifies stereotype threat (Steele & Aronson, 1995) as an influence analogous to his X-factors. Wicherts et al. (2005) found that the presence of experimentally induced stereotype threat led to measurement non-invariance between the experimental and control groups. The non-invariance was easy to detect in a multiple-group confirmatory factor analysis even when sample sizes were modest ($N < 100$).

As Sackett et al. (2004) point out, when stereotype threat was absent in the control condition of Steele and Aronson's study, the IQs of the black and white college students participating in the experiment were what one would have expected them to be on the basis of their prior SAT scores. This indicates that rather than causing the black-white IQ gap, stereotype threat

²⁰ There are some reliable black-white ability differences independently of *g* (e.g., Jensen & Reynolds, 1982), but these are very minor compared to the general differences associated with *g*.

widens a pre-existing gap that is persistently observed regardless of social context. Stereotype threat appears to be no more than yet another curiosity of the psychological laboratory without real-world implications (Lee, 2009). Given the assumed similarity of stereotype threat and Kaplan's X-factors, we would expect the latter to cause measurement non-invariance, too, something that has not been observed in analyses of white and black test scores in non-experimental settings. This strongly suggests that Kaplan's X-factors are non-existent.

5.2.3. Flynn effect

Kaplan believes that the Flynn effect presents another environmental influence supporting his thesis. He claims that there is "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." It is true that the Flynn effect is not associated with increased test score variances, underlining the inadequacy of Kaplan's variance difference tests as a way to discover environmental influences. However, it is not true that more appropriate methods fail to detect the Flynn effect.

When tests of measurement invariance, described above, and analogous differential item functioning tests have been applied to IQ data from different age cohorts, it has consistently been found that measurement invariance between cohorts is untenable (Wicherts et al., 2004; Beaujean & Osterlind, 2008; Must et al., 2009; Wai & Putallaz, 2011; Shiu et al., 2013; Pietschnig et al., 2013; Fox & Mitchum, 2013, 2014; Beaujean & Sheng, 2014). As Wicherts et al. (2004) point out, the fact that black-white IQ differences are associated with measurement invariance while the Flynn effect is not indicates that the two phenomena are separate, and that one of them does not tell us anything about the other. Consistently with this finding, Ang et al. (2010) found that the magnitude of the Flynn effect does not differ between races. The environmental improvements underlying the Flynn effect have reached blacks and whites equally, suggesting that the environmental factors influencing cognitive development are highly similar in the two races.

Contrary to what Kaplan believes, the Flynn effect is easy to identify with standard psychometric methods and ordinary sample sizes. If he wants to maintain that his racial X-factors would not be detectable with the same methods, he must modify his thesis and argue that the influence of his X-factors is uniquely subtle and completely different in character from known environmental influences such as the Flynn effect.

5.2.4. Rowe and colleagues' findings

One of the principal targets of Kaplan's article are two studies by David Rowe and colleagues (1994, 1995). These studies investigated variance-covariance and correlation matrices of "environmental" influences (e.g., quality of child's home environment, mother's education, and parents' school involvement) and outcome variables (e.g., IQ, self-esteem, and delinquency) across different races and ethnic groups. This study design where the equality of matrices is directly compared represents a model-free analogue to the model-based analyses of measurement invariance discussed above (although mean vectors were not examined by Rowe et al.). While the model-based analyses examine the statistical structure of individual differences in IQ test performance, Rowe and colleagues extended the same logic to an analysis of a wide range of variables beyond tests. Both methods rely on the insight that the effects of X-factors will not be limited to a specific variable, but rather will ramify across a whole network of related variables, reorganizing their mutual relations in a way that can be detected with statistical techniques. X-factors are expected to cause differences especially in the covariances (or correlations) of observed variables across groups.

Kaplan's model assumes that "racialized environments" simultaneously reduce IQ and make the environmental circumstances of blacks worse, which should show up as increases in the covariances between IQ and measured environmental factors. Similarly, one would expect Kaplan's X-factors, if they exist, to negatively influence not only the IQs of black children, but also their self-esteem and aspirations, increasing the associations between these variables. In contrast, there is no way to say if the *variance* of IQ scores, which is the only statistic that Kaplan is interested in, should be lower, the same, or higher in blacks due to the influence of X-factors, given that we do not know what the variance would be without the putative influence of the X-factors.

Rowe and colleagues found the many matrices of environmental and outcome variables that they analyzed to be statistically indistinguishable across groups. Therefore, there appear to be no group-specific sources of developmental differences, or X-factors. This corroborates the consistent finding of measurement invariance between races in confirmatory factor analyses of IQ batteries. Group differences in the mean level of IQ can be attributed to differences in developmental antecedents that are common to all groups.

Therefore, black individuals tend to have low IQ scores for the very same reasons that (a smaller proportion of) white individuals have low IQ scores. These reasons plausibly include genetic differences, but if group differences

are to be explained in completely non-genetic terms, then the causes must be VE-type factors: the IQ-decreasing environments experienced by most blacks have to be similar to those experienced only by disadvantaged whites. However, as discussed above, the available empirical evidence argues strongly against the existence of such VE-factors. The task of the non-hereditarian is further complicated by the fact that genetic and environmental factors show differential associations with different cognitive ability parameters, and black IQ deficits closely resemble genetic influences in this respect.

5.3. Can X-factors influence g ?

The prospect of Kaplan's X-factors not being detected in an analysis of measurement invariance is very poor. This is because they present an influence on test scores that is orthogonal to the influence exerted by latent factors, whereas black-white cognitive differences can in fact be attributed to latent factors. In particular, black-white differences on cognitive tests are positively correlated with the g loadings of the tests, and can be mostly explained by a racial difference in the mean level of g . Kaplan's model cannot account for the observed pattern of g -linked differences.

However, there is a theoretical possibility of X-factors causing g -linked black-white gaps and not violating measurement invariance. That would happen if the X-factors directly influenced g , with their effect on observed test scores fully mediated by latent abilities. This would ensure that the X-factor-induced racial gaps could be attributed to the latent abilities (i.e., measurement invariance), and that the gaps would be correlated with g loadings (because X-factors would explain some of the variance in g). A model like this is depicted in Figure 3.

Is it plausible that X-factors would exclusively and directly influence g ? As discussed earlier, it has consistently been found that environmental influences on test performance are negatively or not at all associated with g loadings, whereas genetic influences are associated strongly and positively with g loadings. Unless the nature of the racial X-factors is completely unique in the domain of environmental influences, they would not cause g -linked gaps. Furthermore, as we have seen, the environmental factors that Kaplan offers as analogues to his X-factors do not cause test score gaps that can be attributed to latent abilities—this is true of both the stable, trait-like gaps associated with the Flynn effect, and the ephemeral, state-like gaps associated with stereotype threat.

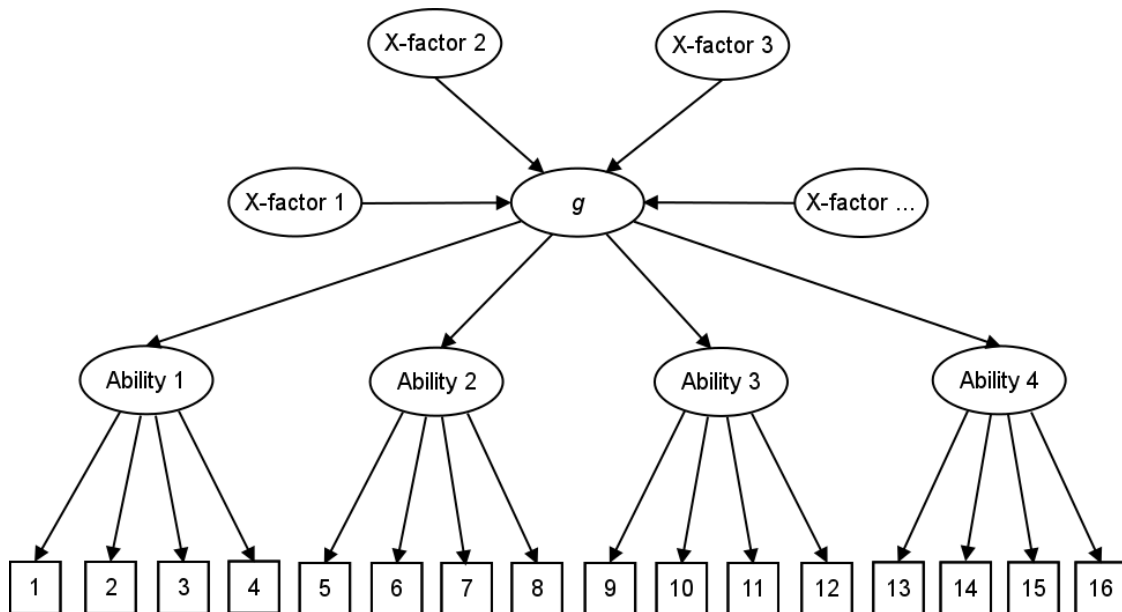


Figure 3. A model where X-factors influence g directly and test scores indirectly. Residual variances are not shown.

While g is overwhelmingly a genetic phenomenon, there are nevertheless some non-genetic influences on it. For example, Panizzon et al. (2014) found that in a large sample of middle-aged male twins the heritability of the latent g factor was 86 percent, with 14 percent accounted for by the non-shared environment. Could X-factors be included in that 14 percent? As discussed above, no environmental factors directly affecting g have been identified, suggesting that environmental influences on g may not have anything to do with aspects of the social environment but rather that they may consist of random, noise-like influences affecting individual development regardless of external circumstances (Kan et al., 2010). Kaplan posits that there is a large number of environmental X-factors, many of them affecting only certain subgroups of blacks, so to assume that all these X-factors would have the same, laser-like focus on g , completely unlike how all known environmental factors influence test scores, makes this model so implausible as to leave it devoid of interest.

7. Non-cognitive differences between blacks and whites

Kaplan's model of the IQ gap presupposes that the daily lives of African Americans are saturated with racially motivated insults and humiliations that inflict serious psychological trauma on them. The nature of these negative experiences is such that one would expect them to have their most direct and most profound effects on non-cognitive rather than cognitive characteristics. If Kaplan had presented his model as an explanation of racial

differences in the prevalence of some psychiatric disorder rather than in the mean level of IQ, it would have had some prior plausibility, given the well-established link between stressful life events and mental disorders (e.g., Hammen, 2005). “Racialized environments” would be expected to discourage blacks in their pursuits, lower their self-esteem, and lead to a high prevalence of mood disorders such as depression and social phobia among them. It is difficult to imagine why the emotional well-being, motivation, and self-concept of blacks would not suffer from the same experiences that supposedly greatly harm their cognitive abilities. Therefore, one would expect that if Kaplan’s model were correct, measures of relevant non-cognitive characteristics would show even larger white-black gaps than the ones seen on tests of cognitive ability.

Table 1 lists variables related to emotional well-being, self-confidence, and optimism, broadly construed, with the gaps between whites and blacks on these variables reported in terms of Cohen’s d . The data are from various meta-analyses and large, nationally representative studies, as indicated in the table. They are coded in such a way that a *positive* (>0.00) gap always indicates that, on the average, blacks are better off on the particular variable than whites, while a *negative* (<0.00) gap indicates that whites are, on the average, better off. For example, the self-esteem gap of $+0.19$ means that blacks tend to have higher self-esteem, and the panic disorder gap of $+0.28$ means that the disorder is more common in whites, while the bipolar disorder gap of -0.10 indicates that this disorder is more common in blacks. For comparative purposes, the black-white IQ gap is also presented in the table.

Two things are immediately evident from Table 1. Firstly, there are no racial differences in the non-cognitive variables that could be characterized as large or even medium-sized in terms of Cohen’s (1988) taxonomy of effect sizes. The differences are small to very small, providing a stark contrast to the IQ gap which stands at $d = -1.10$, representing a very large effect. Secondly, blacks appear to suffer from many psychiatric disorders somewhat less frequently than whites, and they generally have at least as optimistic and confident an outlook on life as whites.

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

Variable	Black-white difference (<i>d</i>)	Reference	Notes
Self-esteem	0.19	Twenge & Crocker (2002)	Meta-analysis, K=354
Positive affect	0.05	original analysis	Analysis of public-use data from Wave IV of the Add Health study; representative national sample, ages 24–34; positive affect operationalized as in De Neve & Oswald (2012); variables used: H4MH20, H4MH24, H4MH25, and GSWG4_2 (weights)
Educational aspirations	-0.01	Mau & Bikos (2000)	Representative national sample of adolescents and young adults (males and females pooled)
Occupational aspirations	0.00	"	"
Panic disorder	0.28	Breslau, et al. (2006)	Representative national sample, odds ratio for lifetime risk converted to <i>d</i>
Generalized anxiety disorder	0.28	"	"
Social phobia	0.12	"	"
Post-traumatic stress disorder	0.00	"	"
Bipolar disorder	-0.10	"	"
Dysthymic disorder	-0.13	Riolo et al. (2005)	Representative national sample, age 15–40, unadjusted odds ratio converted to <i>d</i>
Major depressive disorder	0.27	"	"
Female body dissatisfaction	0.33	Roberts et al. (2006)	Meta-analysis of “global” body dissatisfaction, K=38
Suicidal thoughts or attempts	0.15	Blum et al. (2000)	Representative national sample of 7th to 12th graders, odds ratio converted to <i>d</i>
Suicide rate	0.43	Williams & Jackson (2005)	Death from suicide in 2000, National Center for Health Statistics data, odds ratio converted to <i>d</i>
IQ	-1.10	Roth et al. (2001)	Meta-analysis of racial differences in general cognitive ability, K=105

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

Several objections can be presented against these results. It is possible that due to lack of access to health care, blacks are underdiagnosed with respect to the disorders examined here. Similarly, the self-report measures used could reflect a greater tendency towards socially desirable responding in blacks, and black suicides may remain unidentified more often than white ones. However, the reported gaps generally favor blacks, and if the true effect sizes really had the opposite sign and were large, the bias in the measures used would have to be implausibly pervasive. For example, if the real gap in social phobia, obscured by underdiagnosis, were -1.10 (favoring whites), instead of the actually observed gap of 0.12 (favoring blacks), it would mean that the real lifetime prevalence of the disorder is more than 50 percent in blacks, compared to the observed prevalence of 10.8 percent, assuming that there is no underdiagnosis at all in whites whose observed lifetime prevalence is 12.6 percent. We can safely conclude that the basic pattern of results shown in Table 1 does not stem from measurement bias.

Another objection might be that the levels of emotional well-being and self-esteem in blacks as compared to whites could be inherently higher (for genetic or cultural reasons), so that even very traumatic experiences would not altogether eliminate the black advantage. However, this explanation is entirely *ad hoc* and without evidentiary basis, and it is contrary to the tenor of Kaplan's argument (so he would probably not endorse it).²¹ An even less promising conjecture to explain the results in Table 1 is that racism has *beneficial* effects on blacks, making them strive more to prove themselves and providing protection against mental ailments, while simultaneously causing large cognitive and academic deficits in them.

All in all, these results present a strong disconfirmation of Kaplan's model, corroborating the psychometric evidence against the model presented in previous sections. The personal characteristics that one would expect to be most directly and potently affected by the kind of chronically racially biased society that Kaplan describes are in fact generally not affected at all. On the contrary, the data show African Americans to be at least as well-adjusted as whites. Black Americans appear to possess a great deal of confidence in their abilities and a very optimistic attitude to life, as exemplified by the fact that the educational and occupational aspirations of black adolescents and young adults are virtually identical to those of their white peers, in spite of the large white advantage in academic performance.²² Kaplan's portrayal of

²¹ I would argue that the differences presented in Table 1 may reflect both genetic and environmental causes, but that they are unlikely to have anything to do with racism.

²² Mello (2009) found that the educational and occupational aspirations of blacks are in fact consistently higher than those of whites. The difference to Mau & Bikos (2000),

black Americans as psychologically traumatized victims of a racist society is bluntly contradicted by these findings.

8. How to test HM

Kaplan argues that HM is not a testable scientific proposition. This claim is mistaken. As previously noted by Rowe (2005), Murray (2005), Rushton & Jensen (2005, p. 262), and Lee (2009), among others, one of the appealing features of HM is that there exists an *experimentum crucis* whose outcome could settle the issue once and for all. This natural experiment is fully feasible using current technology.

The study design would exploit the fact that African Americans are an admixed population with a major West African and a smaller European element, while white Americans are almost exclusively descended from European immigrants (Lao et al., 2010). On the average, the ancestry of black Americans is approximately 80 percent African and 20 percent European, but, crucially, these percentages vary considerably across individuals—the standard deviation is about 12 percentage points (Bhatia et al., 2013). Modern genomic methods using ancestrally informative markers enable the accurate partition of an individual's ancestry into African, European, and other ancestral components (Kosoy et al., 2009). Because genetic influence on IQ mostly reflects the additive effects of up to thousands of genes (Davies et al., 2011), HM predicts that there is a strongly positive and linear relation between IQ and the extent of white ancestry in African Americans. In other words, a greater amount of white admixture is assumed to bring with it a more advantageous mix of alleles influencing IQ.

Therefore, one only needs to recruit a large, representative sample of black Americans and obtain from each of them a valid IQ score and a DNA-based estimate of European admixture. If HM is correct, there should be a strong correlation between white ancestry and IQ. To ensure that any possible association is not driven by correlations between ancestry and physical appearance, appropriate covariates (e.g., skin color) can be used in the analysis. The most direct and powerful way of ruling out the influence of confounding variables would be to use a sibling fixed effects design where IQ and ancestry are investigated within sibling pairs.

whose results are presented in Table 1, probably reflects the fact that Mello used more comprehensive measures of aspirations than Mau and Bikos. Mello's results were not included in Table 1 because appropriate *d* values cannot be straightforwardly calculated from her paper. The discrepancy between aspirations and achievement among blacks is a well-established and long-standing phenomenon, often called the *attitude-achievement paradox* (Downey et al., 2009).

This admixture design has been frequently used in biomedical research. The degree of African ancestry has been found to be associated with, for example, preterm birth (Tsai et al., 2011), osteoporosis (Chen et al., 2011), body mass index (Nassir et al., 2012), diabetes (Cheng et al., 2012), asthma (Flores et al., 2012), and hypertension (Kosoy et al., 2012). Of greater interest to the present discussion is the finding that African ancestry is negatively correlated with educational and occupational attainment and family income in black Americans (Cheng et al., 2012, Table S2). This finding greatly complicates theories that attribute the black-white IQ gap to social class differences.

The feasibility of admixture analysis means, at the very least, that HM is *falsifiable*. If no correlation between IQ and ancestry were found in African Americans, HM would have to be rejected, and a redoubled effort at identifying environmental causes of racial differences could commence. In contrast, if white ancestry were found to be strongly associated with greater IQ, it would provide very powerful evidence in favor of HM, but I would expect that many committed anti-hereditarians would still not accept HM. Even so, a high correlation between ancestry and IQ would necessarily greatly constrain many proposed models of environmental causation. For example, Kaplan's theory of racialized environments would have to be modified to accommodate the notion that the effects of racism on IQ are heavily moderated by largely cryptic differences in ancestry. Considering that few black Americans have knowledge of their precise ancestry, it would be very challenging to explain high IQ-ancestry correlations in purely social terms.

Thus, contrary to Kaplan's claim that "given the actual state of the world there is no way to generate any reasonably strong evidence in favor of the hereditarian hypothesis", HM is an eminently testable scientific model. In contrast, the non-hereditarian explanation of the black-white IQ gap is essentially unfalsifiable because even in the face of overwhelming evidence in favor of HM, it is always possible to postulate that some exotic and imperceptible environmental influence is to blame for the gap.²³

9. Discussion

James Flynn has criticized researchers for assuming that racism is a magical ambient force rather than one whose possible effects are manifested through such ordinary mechanisms as poverty and poor self-esteem. Kaplan

²³ Kaplan's suggestion that transgenerational epigenetic influences may be implicated in the black-white gap is an example of the kind of wildly implausible mechanism that can be posited to save the non-hereditarian model from falsification.

rejects this argument. Indeed, Kaplan's racial X-factors resemble nothing so much as magic. He presents no evidence for the hypothesis that what he calls racialized environments have an effect on IQ, and his evidence for the very existence of these environments is very weak. Nevertheless, his model presupposes that such environments, no matter how heterogeneous, act like magic bullets, causing large, *g*-linked cognitive deficits in blacks from all backgrounds while miraculously bypassing all the brain systems that mediate emotional and motivational processes. Furthermore, the racial X-factors do all this in such a subtle way that no statistical signals of their presence can ever be observed, making racism a causal force completely unlike all known environmental influences on IQ scores. The essentially occult powers that Kaplan attributes to white racism take his arguments beyond the bounds of science.

A fundamental flaw in Kaplan's thesis is that of the many lines of evidence presented by hereditarians, he considers only one, Jensen's binary of VE-factors and X-factors. Thinking that he has refuted this particular argument, Kaplan concludes that HM as a whole is untenable. However, HM consists of a large body of interlocking theoretical arguments and pieces of empirical evidence (not all of which have been explicitly considered in this article) which should not be investigated in isolation from each other. Postulating X-factors to explain the IQ gap is an empty exercise unless one shows that such factors fit the totality of evidence. Because Kaplan fails to consider all the relevant facts, his X-factor model could be correct only if a long list of assumptions that he leaves unstated and unexplored were correct. When those assumptions are spelled out, the model's fatal flaws come into view.

The fact that Kaplan's proposed X-factors turn out to be very elusive upon closer inspection attests to the wisdom of Jensen's argument about the non-existence of X-factors in general. Considering that Kaplan is an associate professor of philosophy, another lesson that might be drawn from his very confidently presented yet completely unsuccessful challenge to HM is that a philosophical education alone is without value in a scientific dispute. A good command of the theories, methods, and evidence pertinent to the particular area of research is necessary for making useful scientific contributions.

Kaplan attacks HM on rather general grounds and appears to be largely ignorant of the extensive network of evidence that makes HM such a compelling model. In particular, the psychometric aspects of Kaplan's model are so underdeveloped that it cannot be properly tested by simulation, but he nevertheless thinks that his simulations provide strong evidence against HM. If a properly elaborated version of Kaplan's simulation model were put to test in a multiple-group confirmatory factor analysis framework, the chances of his X-factors going undetected would be very small.

Kaplan is also oblivious to the fact that, perhaps uniquely among all the long-running disputes in social science, a definitive empirical resolution to the black-white IQ controversy is within the reach of contemporary science. The strong causal implications of DNA-based admixture studies have been frequently discussed in the literature, and the ability gap between blacks and whites is widely recognized as one of the most significant social problems in America (Jencks & Phillips, 1998; Paige & Witty, 2010; Giles, 2011). That there nevertheless has been no rush to use genomic methods to clarify the etiology of the gap testifies to the taboo nature of the hereditarian model.²⁴ However, continuous progress is being made in elucidating the molecular genetic basis of intelligence (Rietveld et al., 2013; Ward et al., 2014), so we will eventually find the answer anyway.

²⁴ Charles Murray once tried to launch a “bipartisan” effort to investigate the black-white gap with genomic admixture methods, inviting researchers associated with the hereditarian and the non-hereditarian “camps” to design and conduct such a study together, but the project fell through because the non-hereditarians were unwilling to pursue this line of research. See: <https://groups.yahoo.com/neo/groups/evolutionary-psychology/conversations/messages/38171>.

References

- Anacker, K. B., Carr, J. H., & Pradhan, A. (2012). Analyzing foreclosures among high-income Black/African American and Hispanic/Latino borrowers in Prince George's County, Maryland. *Housing and Society, 39*, 1–28
- Ang, S., Rodgers, J. L., & Wänström, L. (2010). The Flynn effect within subgroups in the U.S.: Gender, race, income, education, and urbanization differences in the NLSY-Children data. *Intelligence, 38*, 367–384.
- Beaujean, A. A., & Osterlind, S. J. (2008). Using item response theory to assess the Flynn effect in the National Longitudinal Study of Youth 79 children and young adults data. *Intelligence, 36*, 455–463.
- Beaujean, A. A., & Sheng, Y. (2014). Assessing the Flynn effect in the Wechsler scales. *Journal of Individual Differences, 35*, 63–78.
- Beaver, K. M., DeLisi, M., Wright, J. P., Boutwell, B. B., Barnes, J. C., & Vaughn, M. G. (2013). No evidence of racial discrimination in criminal justice processing: Results from the National Longitudinal Study of Adolescent Health. *Personality and Individual Differences, 55*, 29–34.
- Bhatia, G., et al. (2013). Genome-wide scan of 29,141 African Americans finds no evidence of selection since admixture (Preprint). Retrieved from arXiv: <http://arxiv.org/pdf/1312.2675.pdf>.
- Blum, R. W., Beuhring, T., Shew, M., et al. (2000). The effects of race/ethnicity, income, and family structure on adolescent risk behaviors. *American Journal of Public Health, 90*, 1879–1884.
- Breslau, J., Aguilar-Gaxiola, S., Kendler, K., Su, M., Williams, D., & Kessler, R. (2012). Specifying race-ethnic differences in risk for psychiatric disorder in a USA national sample. *Psychological Medicine, 35*, 1–12.
- Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research*. New York, NY: The Guilford Press.
- Card, D., & Rothstein, J. (2007). Racial Segregation and the Black-White Test Score Gap. *Journal of Public Economics, 91*, 2158–2184.
- Carneiro, P., Heckman, J. J., & Masterov, D. V. (2005). Labor market discrimination and racial differences in pre-market factors. *Journal of Law and Economics, 47*, 1–39.
- Case, P. (2013). Questioning Assumptions about Race, Social Class and Crime Portrayal: An Analysis of Ten Years of Law and Order. *International Journal of Criminology and Sociology, 2*, 240–256.
- Chen, F. F., Sousa, K. H., & West, S. G. (2005). Testing measurement invariance of second-order factor models. *Structural Equation Modeling, 12*, 471–492.

- Chen, Z., Qi, L., Beck, T. J., et al. (2011). Stronger bone correlates with African admixture in African-American women. *Journal of Bone and Mineral Research*, *26*, 2307–2316.
- Cheng, C.-Y., et al. (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.
- Chiricos, T., & Eschholz, S. (2002). The racial and ethnic typification of crime and the criminal typification of race and ethnicity in local television news. *Journal of Research in Crime and Delinquency*, *39*, 400–420.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Coviello, D., & Persico, N. (2013). *An Economic Analysis of Black-White Disparities in NYPD's Stop and Frisk Program* (NBER Working Paper No. 18803). Retrieved from the National Bureau of Economic Research website: <http://www.nber.org/papers/w18803>.
- Currie, J. (2005). Health disparities and gaps in school readiness. *Future of Children*, *15*, 117–138.
- Davies, G., et al. (2011). Genome-wide association studies establish that human intelligence is highly heritable and polygenic. *Molecular Psychiatry*, *16*, 996–1005.
- De Neve, J. E., & Oswald, A. J. (2012). Estimating the influence of life satisfaction and positive affect on later income using sibling fixed effects. *Proceedings of the National Academy of Sciences*, *109*, 19953–19958.
- Deary, I. J. (2012). Intelligence. *Annual Review of Psychology*, *63*, 453–482.
- Deutsch, S.K., & Cavendar, G. (2008). CSI and forensic realism. *Journal of Criminal Justice and Popular Culture*, *15*, 34–53.
- Dixon, T. L., & Linz, D. (2000). Overrepresentation and Underrepresentation of African Americans and Latinos as Lawbreakers on Television News. *Journal of Communication*, *50*, 131–54.
- Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, *35*, 21–50.
- Dolan, C. V., & Hamaker, E. L. (2001). Investigating Black-White differences in psychometric IQ: Multi-group confirmatory factor analyses of the WISC-R and K-ABC and a critique of the method of correlated vectors. In Frank Columbus (Ed.), *Advances in Psychology Research*, vol. 6 (pp. 31–59). Huntington, NY: Nova Science Publishers.
- Downey, D. B., Ainsworth, J. W., & Zhenchao, Q. (2009). Rethinking the Attitude-Achievement Paradox among Blacks. *Sociology of Education*, *82*, 1–19.
- Eschholz, S. (2002). Racial composition of television offenders and viewers' fear of crime. *Critical Criminology*, *11*, 1–20.

- Eschholz, S., Mallard, M., & Flynn, S. (2004). Image of prime time justice: A content analysis of "NYPD Blue" and "Law and Order." *Journal of Criminal Justice and Popular Culture*, *10*, 161–180.
- Espenshade, T. J., & Radford, A. W. (2009). *No Longer Separate, Not Yet Equal: Race and Class in Elite College Admission and Campus Life*. Princeton: Princeton University Press.
- Ferguson, M. F., & Peters, S. R. (1995). What Constitutes Evidence of Discrimination in Lending? *The Journal of Finance*, *50*, 739–748.
- Fox, M. C., Mitchum, A. L. (2013). A knowledge-based theory of rising scores on "culture-free" tests. *Journal of Experimental Psychology: General*, *142*, 979–1000.
- Fox, M. C., Mitchum, A. L. (2014) Confirming the Cognition of Rising Scores: Fox and Mitchum (2013) Predicts Violations of Measurement Invariance in Series Completion between Age-Matched Cohorts. *PLOS ONE* *9*(5): e95780.
- Flores, C., Ma, S. F., Pino-Yanes, M., Wade, M. S., Perez-Mendez, L., Kittles, R. A., et al. (2012). African ancestry is associated with asthma risk in African Americans. *PLoS One* *7*(1):e26807.
- Flynn, J. R. (1980). *Race, IQ and Jensen*. London, UK: Routledge.
- 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.
- Giles, J. (2011). Social science lines up its biggest challenges. *Nature*, *470*, 18–19.
- Gilliam, F. D., Jr., Iyengar, S., Simon, A., & Wright, O. (1996). Crime in Black and White: The Violent, Scary World of Local News. *The Harvard International Journal of Press/Politics*, *1*, 6–23.
- Gottfredson, L. S. (2002). *g*: Highly general and highly practical. in R. J. Sternberg, & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 331–380). Mahwah, NJ: Erlbaum.
- Hammen, C. (2005). Stress and depression. *Annual Review of Clinical Psychology*, *1*, 293–319.
- Heckman, J. J. (1998). Detecting Discrimination. *Journal of Economic Perspectives*, *12*, 101–116.
- Heckman, J. J., Lochner, L. J., & Todd, P. E. (2006). Earnings functions, rates of return and treatment effects: The Mincer equation and beyond. In E. A. Hanushek, & F. Welch (Eds.), *Handbook of the Economics of Education* (pp. 307–458). Amsterdam, NL: North Holland.
- Herrnstein, R. J., & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York, NY: Free Press.

- Jencks, C., & Phillips, M. (1998). The Black-White Test Score Gap: An Introduction. In C. Jencks, & M. Phillips (Eds.), *The Black-White Test Score Gap* (pp. 1–51). Washington, DC: Brookings.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. Westport, CT: Praeger.
- Jensen, A. R., & Reynolds, C. R. (1982). Race, social class and ability patterns on the WISC-R. *Personality and Individual Differences*, 3, 423–438.
- Johnson, W. R., & Neal, D. (1998). Basic Skills and the Black-White Earnings Gaps. In C. Jencks, & M. Phillips (Eds.), *The Black-White Test Score Gap* (pp. 480–497). Washington, DC: Brookings.
- Kalev, A., Dobbin, F., & Kelly, E. (2006). Best Practices or Best Guesses? Assessing the Efficacy of Corporate Affirmative Action and Diversity Policies. *American Sociological Review*, 71, 589–617.
- Kan, K.-J., Ploeger, A., Raijmakers, M. E. J., Dolan, C. V., & van der Maas H. L. J. (2010). Nonlinear epigenetic variance: review and simulations. *Developmental Science*, 13, 11–27.
- Kaplan, J. M. (2014). Race, IQ, and the search for statistical signals associated with so-called “X”-factors: environments, racism, and the “hereditarian hypothesis”. *Biology & Philosophy*. Advance online publication.
- Kendler, K. S., & Baker, J. H. (2007). Genetic influences on measures of the environment: A systematic review. *Psychological Medicine*, 37, 615–626.
- Kosoy, R., Nassir, R., Tian, C., et al. (2009). Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. *Human Mutation*, 30, 69–78.
- Kosoy, R., Qi, L., Nassir, R., et al. (2012). Relationship between hypertension and admixture in post-menopausal African American and Hispanic American women. *Journal of Human Hypertension*, 26, 365–373.
- Laderman, E., & Reid, C. (2008). *Lending in Low- and Moderate-Income Neighborhoods in California: The Performance of CRA Lending During the Subprime Meltdown* (Federal Reserve Bank of San Francisco Working Paper 2008-05). Retrieved from the Federal Reserve Bank of San Francisco website: <http://www.frbsf.org/economic-research/files/wp08-05.pdf>.
- Lao, O., Vallone, P. M., Coble, M. D., Diegoli, T. M., van Oven, M., et al. (2010). Evaluating self-declared ancestry of U.S. Americans with autosomal, Y-chromosomal and mitochondrial DNA. *Human Mutation*, 31, E1875–E1893.
- Lee, J. J. (2009). Review of Intelligence and How to Get It: Why Schools and Cultures Count. *Personality and Individual Differences*, 48, 247–255.
- Lubke, G. H., Dolan, C. V., Kelderman, H., & Mellenbergh, G. J. (2003). On the relationship between sources of within- and between-group differences

- and measurement invariance in the context of the common factor model. *Intelligence*, *31*, 543–566.
- MacDaniel, M. A., & Kepes, S. (2012). *Spearman's Hypothesis Is a Model for Understanding Alternative g Tests*. Paper presented at the 27th Annual Conference of the Society for Industrial and Organizational Psychology, San Diego, CA.
- Mau, W.-C., & Bikos, L. H. (2000). Educational and Vocational Aspirations of Minority and Female Students: A Longitudinal Study. *Journal of Counseling & Development*, *78*, 186–194.
- Mello, Z. R. (2009). Racial/ethnic group and socioeconomic status variation in educational and occupational expectations from adolescence to adulthood. *Journal of Applied Developmental Psychology*, *30*, 494–504.
- Murray, C. (2005). The Inequality Taboo. *Commentary*, *120*, 13–22.
- 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.
- Nassir, R., Qi, L., Kosoy, R., et al. (2012). Relationship between adiposity and admixture in African-American and Hispanic-American women. *International Journal of Obesity*, *36*, 304–313.
- New Century Foundation (2005). *The Color of Crime. Race, Crime and Justice in America* (2nd ed.). Oakton, VA: Author.
- Office of Federal Contract Compliance Programs (2002). *Facts on Executive Order 11246—Affirmative Action*. Retrieved from <http://www.dol.gov/ofccp/regs/compliance/aa.htm>.
- Pager, D., & Shepherd, H. (2008). The Sociology of Discrimination: Racial Discrimination in Employment, Housing, Credit, and Consumer Markets. *Annual Review of Sociology*, *34*, 181–209.
- Paige, R., & Witty, E. (2010). *The Black-White Achievement Gap: Why Closing It is the Greatest Civil Rights Issue of Our Time*. New York, NY: American Management Association.
- Panizzon, M. S., et al. (2014). Genetic and environmental influences on general cognitive ability: Is *g* a valid latent construct? *Intelligence*, *43*, 65–76.
- Phillips, M., et al. (1998). Family background, parenting practices, and the black–white test score gap. In C. Jencks, & M. Phillips (Eds.), *The Black–White Test Score Gap* (pp. 103–144). Washington, DC: Brookings.
- Pietschnig, J., Tran, U. S., & Voracek, M. (2013). Item-response theory modeling of IQ gains (the Flynn effect) on crystallized intelligence: Rodgers' hypothesis yes, Brand's hypothesis perhaps. *Intelligence*, *41*, 791–801.

- Plomin, R., Reiss, D., Hetherington, E. M., & Howe, G. (1994). Nature and nurture: Genetic contributions to measures of the family environment. *Developmental Psychology, 30*, 32–43.
- Plomin, R., & Spinath, F. M. (2004). Intelligence: genetics, genes, and genomics. *Journal of Personality and Social Psychology, 86*, 112–129.
- Potter, W. J., Vaughan, M. W., Warren, R., Howley, K., Land, A., & Hagemeyer, J. C. (1995). How Real is the Portrayal of Aggression in Television Entertainment Programming? *Journal of Broadcast Electronic Media, 39*, 179–192.
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., & Martin, N. W. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science, 340*, 1467–1471.
- Riolo, S. A., Nguyen, T. A., Greden, J. F., & King, C. A. (2005). Prevalence of depression by race/ethnicity: findings from the National Health and Nutrition Examination Survey III. *American Journal of Public Health, 95*, 998–1000.
- Roberts, A., Cash, T. F., Feingold, A., Johnson, B. T. (2006). Are black-white differences in females' body dissatisfaction decreasing? A meta-analytic review. *Journal of Consulting and Clinical Psychology, 74*, 1121–1131.
- Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: a meta-analysis. *Personnel Psychology, 54*, 297–330.
- Roth, P. L., Huffcutt, A. I., & Bobko, P. (2003). Ethnic group differences in measures of job performance: a new meta-analysis. *Journal of Applied Psychology, 88*, 694–706.
- Rowe, D. C. (2005). Under the Skin: On the Impartial Treatment of Genetic and Environmental Hypothesis of Racial Differences. *American Psychologist, 60*, 60–70.
- Rowe, D. C., Vazsonyi, A. T., & Flannery, D. J. (1994). No more than skin deep: ethnic and racial similarity in developmental process. *Psychological Review, 101*, 396–413.
- Rowe, D. C., Vazsonyi, A. T., & Flannery, D. J. (1995). Ethnic and racial similarity in developmental process: a study of academic achievement. *Psychological Science, 6*, 33–38.
- Rowe, D. C., Vesterdal, W. J., & Rodgers, J. L. (1998). Herrnstein's syllogism: Genetic and shared environmental influences on IQ, education, and income. *Intelligence, 26*, 405–423.
- Rushton, J. P., & Ankney, C. D. (2009). Whole brain size and general mental ability: A review. *International Journal of Neuroscience, 119*, 692–732.

- Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on Black-White differences in cognitive ability. *Psychology, Public Policy and Law*, *11*, 235–294.
- Rushton, J. P., & Jensen, A. R. (2010). Editorial. 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.
- Sackett, P. R., Hardison, C. M., & Cullen, M. J. (2004). On interpreting stereotypic threat as accounting for African American–White difference in cognitive tests. *American Psychologist*, *59*, 7–13.
- Sanandaji, T. (2009). *Reversion to the Racial Mean and Mortgage Discrimination* (IFN Working Paper No. 811). Retrieved from the Research Institute of Industrial Economics website: http://www.ifn.se/eng/publications/wp/2009_1/811.
- Sesardic, N. (2005). *Making Sense of Heritability*. Cambridge, UK: Cambridge University Press.
- Shiu, W., et al. (2013). An item-level examination of the Flynn effect on the National Intelligence Test in Estonia. *Intelligence*, *41*, 770–779.
- Smit, D. J., et al. (2010). Heritability of head size in Dutch and Australian twin families at ages 0–50 years. *Twin Research and Human Genetics*, *13*, 370–380.
- Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, *69*, 797–811.
- te Nijenhuis, J., Jongeneel-Grimen, B., & Kirkegaard, E. (2014a). Are Headstart gains on the *g* factor? A meta-analysis. *Intelligence*, *46*, 209–215.
- te Nijenhuis, J., Kura, K., & Hur, Y.-M. (2014b). The correlation between *g* loadings and heritability in Japan: A meta-analysis. *Intelligence*, *46*, 275–282.
- te Nijenhuis, J., & van der Flier, H. (2013). Is the Flynn effect on *g*? A meta-analysis. *Intelligence*, *41*, 802–807.
- te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007) Score gains on *g*-loaded tests: No *g*. *Intelligence*, *35*, 283–300.
- Trundt, K. M. (2013). *Construct Bias in the Differential Ability Scales, Second Edition (DAS-II): A comparison among African American, Asian, Hispanic, and White Ethnic Groups*. Unpublished doctoral dissertation, University of Texas, Austin, TX.
- Trzaskowski, M., Dale, P. S., & Plomin, R. (2013a). No genetic influence for childhood behavior problems from DNA analysis. *Journal of the American Academy of Child and Adolescent Psychiatry*, *52*, 1048–1056.

- Trzaskowski, M., Davis, O. S. P., DeFries, J. C., Yang, J., Visscher, P. M., & Plomin, R. (2013b). DNA evidence for strong genome-wide pleiotropy of cognitive and learning abilities. *Behavior Genetics*, *43*, 267–273.
- Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., et al. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence*, *42*, 83–88.
- Tsai, H.-J., et al. (2011). Role of African Ancestry and Gene-Environment Interactions in Predicting Preterm Birth. *Obstetrics & Gynecology*, *118*, 1081–1089.
- Turkheimer, E. (2000). Three laws of behavior genetics and what they mean. *Current Directions in Psychological Science*, *9*, 160–164.
- Turkheimer, E., & Waldron, M. (2000). Nonshared environment: A theoretical, methodological, and quantitative review. *Psychological Bulletin*, *126*, 78–108.
- Twenge, J. M., & Crocker, J. (2002). Race and self-esteem: Meta-analyses comparing Whites, Blacks, Hispanics, Asians, and American Indians and comment on Gray-Little and Hafdahl (2000). *Psychological Bulletin*, *128*, 371–408.
- van der Maas, H. L. J., Kan, K.-J., & Borsboom, D. (2014). Intelligence Is What the Intelligence Test Measures. Seriously. *Journal of Intelligence*, *2*, 12–15.
- Vinkhuyzen, A. A. E., van der Sluis, S., de Geus, E. J. C., Boomsma, D. I., & Posthuma, D. (2010). Genetic influences on 'environmental' factors. *Genes, Brain, and Behavior*, *9*, 276–287.
- Ward, M. E., McMahon, G., St Pourcain, B., Evans, D. M., Rietveld, C. A., et al. (2014). Genetic Variation Associated with Differential Educational Attainment in Adults Has Anticipated Associations with School Performance in Children. *PLoS ONE* *9*(7): e100248.
- Wax, A. (2011). Disparate Impact Realism. *William & Mary Law Review*, *53*, 621–712.
- Why Family Income Differences Don't Explain the Racial Gap in SAT Scores (2008). *The Journal of Blacks in Higher Education*, *62*, 10–12.
- Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice*, *29*, 39–47.
- Wicherts, J., Dolan, C., Hessen, D., Oosterveld, P., van Baal, C., Boomsma, D., & Span, M. (2004). Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, *32*, 509–537.
- Williams D. R., & Jackson, P. B. (2005). Social sources of racial disparities in health. *Health Affairs*, *24*, 325–334.