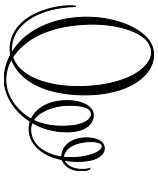
Studying Correlations between Genetic Variation and Test Score Gaps

Studying Correlations between Genetic Variation and Test Score Gaps

Edited by

Gregory Connor and John G.R. Fuerst

Cambridge
Scholars
Publishing



Studying Correlations between Genetic Variation and Test Score Gaps

Edited by Gregory Connor and John G.R. Fuerst

This book first published 2024

Cambridge Scholars Publishing

Lady Stephenson Library, Newcastle upon Tyne, NE6 2PA, UK

British Library Cataloguing in Publication Data
A catalogue record for this book is available from the British Library

ISBN (10): 1-5275-7566-7
ISBN (13): 978-1-5275-7566-0

TABLE OF CONTENTS

INTRODUCTION

GREGORY CONNOR AND JOHN G.R. FUERST

In April 2003, the Human Genome Project announced the successful completion of its project to sequence the three billion base pairs of DNA in the human genome. This landmark achievement sparked an explosion in new genomic technologies, with widespread applications across multiple disciplines. Among a myriad of applications, these new genomic technologies allowed a major step forward in the analysis of genetic variation as a partial cause of intelligence test score gaps between racial and ethnic groups. This volume includes some of the most significant recent contributions to this important research topic. This brief introduction to the volume does not attempt to summarize all the papers herein; it merely provides a birds-eye view and highlights a few key findings shared among them.

The chapters in Part 1 of the book defend an open and uncensored approach to scientific inquiry. This approach is a widely accepted foundation for scientific research in most fields, but unusually controversial in the study of genetic variation and test score gaps. In chapters 1 and 2, Cofnas and Carl each address the question of whether, on balance, research on the possible genetic component of IQ gaps between racial groups should be suppressed. Both these authors employ a utilitarian framework, in this way evaluating whether suppression of scientific findings in this area has positive social value, against the alternative that scientific openness instead might be superior. They both acknowledge that there are potential positive and negative consequences of allowing race and IQ research. Each argues that, on balance, the standard practice of scientific openness is the wiser science policy choice.

Both authors document the extremely harsh, sometimes threatening or even violent, treatment of researchers whose published findings or public statements run contrary to the acceptable public narrative that IQ gaps are entirely environmentally caused. Carl notes that a utilitarian perspective is not universally shared by the opponents of scientific openness; many opponents treat biological identity between races as a "sacred value" and thereby exempt from utilitarian considerations:

> Why does this area of research incite such vitriolic indignation? A likely reason, as Winegard and Winegard (2015) argue, is that for a large number of academics in the West, the notion of biological sameness between groups (classes, sexes, races) has become what Tetlock (2003) calls a "sacred value" (and see Ginges et al., 2017). Sacred values possess at least two important properties. First, they are incommensurable with respect to instrumental values: no amount of sacred value can be traded off for any amount of an instrumental value. And second, proposals to accept such trade-offs are met not merely with rejection, but with moral outrage. Because arguments such as Wilson's, Jensen's and Murray's clearly threaten the sacred value of biological sameness between groups, it is not enough simply to attack the arguments; the defenders of those arguments must be hounded, and their characters impeached. (Carl, Chapter 2 below).

Cofnas reflects on David Reich's warning regarding likely future findings and their implications for current denialism regarding genetic causes of population trait differences:

> I have deep sympathy for the concern that genetic discoveries about differences among populations may be misused to justify racism. But it is precisely because of this sympathy that I am worried that people who deny the possibility of substantial biological differences among populations across a range of traits are digging themselves into an indefensible position, one that will not survive the onslaught of science. In the last couple of decades, most population geneticists have sought to avoid contradicting the orthodoxy. When asked about

> the possibility of biological differences among human populations, we have tended to obfuscate, making mathematical statements in the spirit of Richard Lewontin about the average difference between individuals from within any one population being around six times greater than the average difference between populations … but this carefully worded formulation is deliberately masking the possibility of substantial average differences in biological traits across populations. (Reich 2018, quoted in Cofnas, Chapter 1 below).

Only a few years have passed since Reich wrote those words, but his forecast seems already realized. New findings on genetic causes of test score gaps are statistically strong, robust to model specification, and have been replicated across multiple independent databases. Genetic variation as one source of race-related test score gaps is no longer just a hypothetical possibility, the existing evidence is now quite strong. This change in scientific understanding is due in part to the papers in Part 2 of this volume. This new reality adds urgency to Reich's call to abandon rigid denialism regarding genetic causes of trait variation across groups.

In Chapter 3, Warne lays out a strong empirical case that some test score gaps have partly genetic causes by reviewing five separate lines of supporting evidence. He laments that the politicization of funding opportunities and other politically motivated blockages on this line of research have impeded scientific progress; he issues a call for more open-minded and uncensored research:

> Combined, all five sources of data make a cohesive argument that group differences in intelligence are not fully environmentally caused. I call on psychologists to have an open mind and to investigate the evidence for themselves, starting with the sources I have cited in this article. I also encourage social scientists to make research contributions that can address this question… Controversy thrives in ignorance; it is time to move beyond the basic question of whether genes have an impact on intergroup mean IQs so that scientists can investigate productive questions of the magnitude of this influence and how to accommodate differences in a diverse society. (Warne, Chapter 3 below)

Parts 2 and 3 of the volume are predicated on the open scientific approach advocated in Part 1. Part 2 contains the core scientific material of the volume, with a strong focus on admixture-based analysis of test score gaps.

The distinction between social race, a socially defined identity, and biogeographical ancestry, a genetically determined characteristic, is a crucial component in all the papers in Part 2. Social race refers to an individual's socially defined membership of race or ethnic groups such as Black, white, Hispanic, Native American, Chinese, etc.. This socially defined identity is easily captured, for the vast majority of US individuals, by simply asking each individual to choose one or more from a list of race/ethnicity categories. Biogeographic ancestry (hereafter BGA) denotes an individual's proportional genetic ancestry across a set of continental populations such as African, European, Amerindian, and Asian. Prior to the genetic revolution of the last two decades, these BGA proportions could only be measured very imprecisely. Nowadays, sophisticated new cluster algorithms applied to genotyped DNA samples can identify and accurately measure BGA proportions for most US individuals.

The genotype-based technology to separate SIRE from BGA has sparked off a small scientific revolution in the application of genetic methods to racial identity-related trait variation. Admixture regression involves least-squares regression of a trait on a set of BGA proportions and a set of SIRE variables simultaneously, along with other relevant variables. Admixture regression greatly improves the ability to separate environmental and genetic causes of ethnic and racial group differences in an array of medical, anthropological, and behavioral traits. Admixture regression analysis has been used over the last two decades by genetic epidemiologists and other researchers to study race and ethnicity related differences in alcohol dependence (Zuo et al., 2009), height (Becker et al., 2011), asthma risk (Flores et al., 2012), cardiovascular disease (Bidulescu et al.,

2014), sleep depth (Halder et al., 2015), cigarette smoking behavior (Choquet, Yin, and Jorgenson, 2021), metabolomics (Mehanna et al., 2022), cancer (Rhead et al., 2022), and diabetes (Parcha et al., 2023). The papers in Part 2 are key contributions to a powerful wave of new research applying modern admixture regression methodologies to the investigation of group differences in intelligence. Admixture analysis had long been acknowledged as a potentially powerful method for identifying partly genetic causes of racial and ethnic group differences in intelligence (Shockley, 1973; Nisbett, 2009). Early studies from the 1960s and 1970s had to rely on ancestry estimates based on blood types. These estimates proved unreliable (Reed, 1997) and so more reliable analysis had to wait until admixture proportions could be measured accurately.

In Chapter 4, Warne gives a lucid introduction to the use of admixture regression methods for measuring the contribution of genetic variation to race-related test score gaps. The empirical analysis in Warne's paper is impaired by the small sample size of his database. The use of on-line intelligence tests and self-reported admixture measurement, rather than professionally administered intelligence tests and lab-linked admixture measurement, are other data-related limitations of his study. These data limitations restrict the depth of analysis that it is attainable relative to other recent admixture studies based on large, professionally constructed databases. Despite its data limitations, the paper is a highly recommended introduction to the admixture regression literature. Interestingly, the obvious concern about the replicability of the paper's small-sample findings did not turn out to be a problem – subsequent papers using large, professionally constructed databases reach the same empirical conclusions as Warne did in his modest, exploratory empirical study.

The choices of SIRE categories to include/exclude from the estimation sample, and the choice of BGA clusters, depend upon the available data set and the particularities of the research question. The choices made across the chapters in Part 2 are quite diverse; see Table 1 below and the individual chapters for greater detail.

Table 1: Self-Identified Race and Ethnicity (SIRE) and Biogeographic Ancestry (BGA) Categories Used in the Chapters in Part 2

Chapter	SIRE Categories in Estimation Sample	BGA Categories
4	Black and Hispanic	European, African, Other
5	Black, white, Native American, Asian, Hispanic, Pacific Islander, Multiple-SIRE, and Other SIRE	European, African, Amerindian, East Asian, Central Asian, Oceanian
6	Non-Hispanic Black, non-Hispanic white, and non-Hispanic Black-white biracial	European and African
7	Hispanic, non-Hispanic white, non-Hispanic Black, and non-Hispanic Black-white biracial	European, African, Amerindian
8	Black, white, Hispanic, Native American, East Asian, South Asian, and Other including Multiple-SIRE	European, African, Amerindian, East Asian, and South Asian
9	Black, white, Hispanic, Native American, East Asian, South Asian, and Other including Multiple-SIRE	European, African, Amerindian, East Asian, and South Asian
10	Black, Hispanic, Native American, and Other	European, Amerindian, African, East Asian, and South Asian
11	Black, white, Hispanic, Native American, and Other	European, African, Amerindian, East Asian, and South Asian

There are various other model specification choices that differ across the chapters. One is the choice between including Black-white biracial SIRE as a separate dummy variable (e.g., Chapter 6)

versus assigning Black-white biracials equal ½-½ fractional weightings on the Black and white SIRE dummy variables (e.g., Chapter 9). Black-white biracial is one of the few multiple-SIRE cases where adding a dedicated dummy variable is practical. It is not feasible to add separate dummy variables for all possible multiple-SIRE combinations since some combinations have very few observations; it is necessary either to drop multiple-SIRE individuals from the estimation sample or to use fractional dummy weightings for them. Another specification choice involves choosing whether Hispanic ethnicity should be included separately as a zero-one dummy in the regression or treated equivalently to other SIRE choices with fractional weighting on Hispanic identity and other racial identities; see Chapter 5 which tries both specification choices and compares them.

A difficult specification choice concerns whether to include socio-economic status (SES) as a separate explanatory variable in the regression. Genetic variation causes SES differences across families, so including SES as an explanatory variable will tend to understate total genetic influence of BGA on test scores. On the other hand, SES can directly influence test scores due to environmental effects arising from a less privileged environment, so excluding it may give rise to missing variable biases in the other estimated coefficients. In many of the chapters the admixture regression is estimated both with and without SES, and the reader is left to choose which seems more informative.

The standard regression estimation algorithm does not allow a set of proportion variables which sum to one for every observation, nor a set of dummy variables such that each observation has unit summed exposure, since together with the intercept such a set of explanatory variables generates a linear singularity in the regression matrix. The standard procedure is to delete one of the proportion variables and/or one of the dummy variables to be used as a base case; the remaining estimated coefficients then capture differences relative to that base case variable. Most of the papers in Part 2 delete the European BGA variable and the white SIRE variable from the admixture regressions since these have the most observations in the sample data and thereby are suitable base cases.

Reviewing all the chapters in Part 2, perhaps the strongest empirical finding is that African ancestry has a substantial and statistically significant negative impact on cognitive ability test scores (see Ch. 4, Table 2; Ch. 5, Table 7, model 2; Ch. 7, Table 2; Ch. 8, Table 2; Ch. 9, Table 5; Ch. 11, Tables 1- 6). Black SIRE and African biogeographic ancestry overlap broadly in this US data, hence when SIRE variables are used alone in a regression (excluding BGA variables) Black SIRE has a negative relationship to test scores. However, this Black SIRE impact on test scores appears to be due mostly to its proxy role for African biogeographic ancestry (compare Chapter 5, Table 7, model 1 versus model 2; or Ch. 8, Table 1 versus Table 2). This does not imply that the direct effect of Black SIRE on test scores is zero, only that it seems too small to separately identify in these moderately-large databases. There is also some evidence, not as uniformly powerful across the chapters, that Amerindian BGA has a negative statistical impact and East Asian BGA possibly has a positive statistical impact on test scores (using European BGA as the base case).

The papers in the volume interpret the relation between biogeographic ancestry and cognitive ability as evidence in support of a partial genetic hypothesis for group differences in test scores, but none of them taken individually control for every possible confounding factor. Rather, following conventional scientific practices, the papers build on one another, addressing weaknesses in prior research, noting current limitations, and proposing directions for further research. As in the case of other behavioral genetic research (e.g., adoption studies, twin studies, and genome-wide association studies) no single analysis is dispositive; the strength of the evidence comes from the accumulated results based on multiple, overlapping research designs.

None of the papers fully answer the question as to what caused the observed differences in genetic variation across ancestries. Some of the papers use evolutionary theory and the history of human development to give theoretical reasons for justifying the tested hypothesis that racial test score gaps have partially genetic causes; see for instance chapters 5 and 6. This is an aspect of the research area that may see substantial progress over coming decades.

Genetic variation across ancestries as a source of trait variation implies that there are genetic variants related to the trait and that their population frequency differs across ancestries. It is not necessary to identify these specific genetic variants to perform admixture regression; the existence of such genetic variants is implied by an observed relationship between admixture and trait. Kirkegaard, et al. (Chapter 5) call this *global admixture analysis*, and they differentiate this from *admixture mapping* in which the researcher attempts to identify specific alleles (or tagged allelles serving as proxies) that covary in frequency with the trait and have measurably different frequencies across one or more BGA's. The papers in this volume mostly involve global admixture, but Chapters 6 and 12 take initial steps toward admixture mapping of the BGA-test score relationship. This involves employing polygenic indices for intelligence or educational attainment (constructed by other researchers) and examining the relationship of their index scores to BGA proportions.

Part 3 of the book looks toward the scientific future. It contains two papers that take a natural next step in this research area: exploring whether test score gaps can be linked to statistical differences in the physical characteristics of the human brain across biogeographic ancestries. Chapter 12 (Fuerst, Shibaev, and Kirkegaard) is a monumental study which tests a wide array of empirical hypotheses about genetic variation, brain size and structure, polygenic indices, and intelligence test scores. Chapter 13 (Kirkegaard and Fuerst) relies on the magnetic resonance image data included in the Adolescent Brain and Cognitive Development (ABCD) database and uses the data to explore brain structure differences across biogeographic ancestries. The paper applies an elastic net statistical framework to build a prediction model of intelligence based on high-dimensional multimodel MRI data. They show that an MRI-based intelligence prediction model estimated using the white-only subsample has substantial predictive power for the Black-only subsample, and that this MRI-based model explains 37% of the observed test score gap between the Black and white subsamples even though it is estimated on the white-only subsample. This is clear preliminary evidence for physically-observable genetically-mediated causes of the Black-white test score gap.

Over fifty years ago, Arthur Jensen (1969) challenged the prevailing orthodoxy that test score gaps between racial groups can be ascribed solely to environmental causes. The environment-only theory which Jensen challenged has maintained its hegemony as the only publicly acceptable theory. Jensen made his case using indirect methods of statistical inference since genomic technology at that time did not allow more direct analysis of genetic data. With the advent of the genomic revolution, quantitative genetic methods including admixture regression have revolutionized research on the causes of test score gaps. The papers in this volume provide a powerful body of evidence in support of Jensen's fifty-year-old claim. None of the papers on its own is dispositive but taken together they constitute a powerful new body of evidence contributing to a better understanding of these recalcitrant test score gaps. With this new genetic evidence, the environment-only theory is less scientifically plausible.

Acknowledgements: All the papers in this volume appeared previously in other publication outlets; we thank the authors and previous publishers for permission to reprint them here. The original publication citations are included in the reference list below.

References

Becker, N. S., Verdu, P., Froment, A., Le Bomin, S., Pagezy, H., Bahuchet, S., and Heyer, E. (2011). Indirect evidence for the genetic determination of short stature in African Pygmies. *American journal of physical anthropology,* 145(3), 390-401.

Bidulescu, A., Choudhry, S., Musani, S. K., Buxbaum, S. G., Liu, J., Rotimi, C. N., … and Gibbons, G. H. (2014). Associations of adiponectin with individual European ancestry in African Americans: the Jackson Heart Study. *Frontiers in genetics*, 5, 22.

Carl, N. (2018) How stifling debate around race, genes and IQ can do harm*, Evolutionary Psychological Science*, 4: 399-407.

Cofnas, N. (2020) Research on group differences in intelligence: A defense of free inquiry, *Philosophical Psychology*, 33:1, 125-147.

Colman, A. M. (2016). Race differences in IQ: Hans Eysenck's contribution to the debate in the light of subsequent research. *Personality and Individual Differences*, 103, 182-189.

Connor, G., and J.G. Fuerst (2022) Linear and partially linear models of behavioral trait variation using admixture regression. In R. Lynn (Ed.). *Intelligence, Race and Sex: Some Controversial Issues: A Tribute to Helmuth Nyborg at 85*. Arktos Publishers.

Choquet, H., Yin, J., and Jorgenson, E. (2021). Cigarette smoking behaviors and the importance of ethnicity and genetic ancestry. *Translational psychiatry*, 11(1), 1-10.

Flores, C., Ma, S. F., Pino-Yanes, M., Wade, M. S., Pérez-Méndez, L., Kittles, R. A., … and Garcia, J. G. (2012). African ancestry is associated with asthma risk in African Americans. *PloS one*, 7(1), e26807.

Fuerst, J. G. (2021). Robustness Analysis of African Genetic Ancestry in Admixture Regression Models of Cognitive Test Scores. *Mankind Quarterly*, 62(2).

Fuerst, J.G.R., E.O.W. Kirkegaard and D. Piffer (2021). More Research Needed: There is a Robust Causal vs. Confounding Problem for Intelligence-associated Polygenic Scores in Context to Admixed American Populations. *Mankind Quarterly* 62(1):151-185.

Fuerst, J.G., E.O. Kirkegaard and V. Shibaev (2023). A Genetic Hypothesis for American Race/Ethnic Differences in Mean g: A Reply to Warne (2021) with Fifteen New Empirical Tests Using the ABCD Dataset. *Mankind Quarterly* 63(4):527-600.

Fuerst, J.G., M. Hu, and G. Connor (2021) Genetic ancestry and general cognitive ability in a sample of American youths, *Mankind Quarterly*, *62:*1, 186–216

Gottfredson, L. S. (2007). Applying double standards to "divisive" ideas. *Perspectives on Psychological Science, 2*(2), 216-220.

Halder, I., Matthews, K. A., Buysse, D. J., Strollo, P. J., Causer, V., Reis, S. E., and Hall, M. H. (2015). African genetic ancestry is associated with sleep depth in older African Americans. *Sleep*, *38*(8), 1185-1193.

Jensen, A.R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, 39, 1-123.

Kirkegaard, E.O.W., & Fuerst, J.G.R. (2023). A multimodal MRI-based predictor of intelligence and its relation to socio-demographic characteristics. *Mankind Quarterly,* 63: 374-397.

Kirkegaard, E.O., R.L. Williams, J.G. Fuerst and G. Meisenberg. (2019). Biogeographic Ancestry, Cognitive Ability and Socioeconomic Outcomes. *Psych*, *1*(1), 1-25.

Lasker, J., B.J. Pesta, J.G. Fuerst and E.O. Kirkegaard (2019), Global ancestry and cognitive ability, *Psych* 1(1): 431-459.

Mehanna, M., McDonough, C. W., Smith, S. M., Gong, Y., Gums, J. G., Chapman, A. B., ... and Cooper-DeHoff, R. M. (2022). Influence of Genetic West African Ancestry on Metabolomics among Hypertensive Patients. *Metabolites*, *12*(9), 783.

Meyer, M. N., Appelbaum, P. S., Benjamin, D. J., Callier, S. L., Comfort, N., Conley, D., ... and Parens, E. (2023). Wrestling with Social and Behavioral Genomics: Risks, Potential Benefits, and Ethical Responsibility. *Hastings Center Report*, *53*, S2-S49.

Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count*. WW Norton and Company.

Parcha, V., Heindl, B., Kalra, R., Bress, A., Rao, S., Pandey, A., ... and Arora, P. (2022). Genetic European ancestry and incident diabetes in Black individuals: Insights from the SPRINT trial. *Circulation: Genomic and Precision Medicine*, *15*(1), e003468.

Reed, T. E. (1997). " The genetic hypothesis": It was not tested but it could have been. *American Psychologist*, 52(1), 77-78.

Rhead, B., Hein, D., Pouliot, Y., Guinney, J., De La Vega, F., and Sanford, N. N. (2022). Genetic ancestry differences in tumor mutation in early and average-onset colorectal cancer.

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

Shibaev, V. and J.R. Fuerst (2023). A Genetically Informed Test of the Cognitive-Colorism Hypothesis. *Natural Systems of Mind*, *3*(1).

Shockley, W. (1973). Variance of Caucasian admixture in Negro populations, pigmentation variability, and IQ. In *Proceeding of the National Academy of Sciences of the United States of America,* 70(7), 2180-2180.

Templeton, A.R. The genetic and evolutionary significance of human races. In *Race and Intelligence: Separating Science from Myth*; Fish, J.M., Ed.; Routledge: Abingdon, UK, 2001.

Warne, R. T. (2020). Continental Genetic Ancestry Source Correlates with Global Cognitive Ability Score. *Mankind Quarterly*, 60(3), 400–422.

Warne, R.T. (2021) Between-group differences in intelligence in the United States are >0% genetically caused: Five converging lines of evidence, *The American Journal of Psychology*, 134: 480-501.

Zuo, L., Luo, X., Listman, J. B., Kranzler, H. R., Wang, S., Anton, R. F., ... and Gelernter, J. (2009). Population admixture modulates risk for alcohol dependence. *Human genetics*, 125(5), 605-613.

PART 1:

SOCIAL AND POLITICAL CONSIDERATIONS IN RESEARCH ON TEST SCORE GAPS

CHAPTER 1

RESEARCH ON GROUP DIFFERENCES IN INTELLIGENCE: A DEFENSE OF FREE INQUIRY[1]

NATHAN COFNAS

Abstract

In a very short time, it is likely that we will identify many of the genetic variants underlying individual differences in intelligence. We should be prepared for the possibility that these variants are not distributed identically among all geographic populations, and that this explains some of the phenotypic differences in measured intelligence among groups. However, some philosophers and scientists believe that we should refrain from conducting research that might demonstrate the (partly) genetic origin of group differences in IQ. Many scholars view academic interest in this topic as inherently morally suspect or even racist. The majority of philosophers and social scientists take it for granted that *all* population differences in intelligence are due to environmental factors. I argue that the widespread practice of ignoring or rejecting research on intelligence differences can have unintended negative consequences. Social policies predicated on environmentalist theories of group differences may fail to achieve their aims. Large swaths of academic work in both the humanities and social sciences assume the truth of environmentalism and are vulnerable to being undermined. We have failed to work through the moral implications of group differences to prepare for the possibility that they will be shown to exist.

Keywords: hereditarianism, environmentalism, utilitarianism, intrinsic value of truth

1. Introduction

The idea that all human groups have, on average, the same potential to develop all psychological traits and abilities is an empirically testable claim, but in some important ways it is unlike most other empirical claims investigated by scientists. Many scientists, philosophers, and even laypeople hold that, for a variety of reasons, we are morally obligated to affirm this proposition, or refrain from conducting research that might undermine it (e.g., Block & Dworkin, 1974; Chomsky, 1976, 1988; Dummett, 1981; Gardner, 2001; Horgan, 2013; Kourany, 2016; Sternberg, 2005). Prominent researchers who publicly endorse hereditarianism about group differences in intelligence[2] have been condemned as immoral, fired from their positions, and physically threatened (examples are documented in Ceci & Williams, 2009; Cofnas, 2016; Gottfredson, 2010; Sesardic, 2005). Many scholars assume that there is no legitimate reason to be interested in group differences. According to Rose (2009), the fact that psychometricians focus on the IQ gap between some groups (e.g.,

[1] Editors' note: We have reformatted the original paper for this collection.

[2] Practically all informed scholars are, to varying degrees, hereditarians about *individual* differences in intelligence (Neisser et al., 1996). The controversy centers on the role of genes in *group* differences. Hereditarians about group differences believe that genes play a nontrivial role in these differences—environmentalists, that genes play no significant role.

Blacks vs. Whites) rather than others (e.g., north vs. south Welsh) "calls into question the motivation behind looking for such specific group differences in intelligence, sheds doubt on whether such research is well-founded, and begs whether answers could possibly be put to good use" (p. 786). He concludes that "in a society in which racism and sexism were absent, the questions of whether whites or men are more or less intelligent than blacks or women would not merely be meaningless—they would not even be asked" (p. 788). Virtually no scientific hypothesis besides race and sex differences is rejected in this way.

The fact that the majority of genetic variation in our species exists within rather than between races, as Lewontin (1972) discovered, does not rule out the possibility of significant race differences in psychological traits like intelligence. Lewontin argued that, because there is more genetic variation within than between races, racial classifications do not correspond to any (genetically based) population structure. That implies that there can be no genetically based racial classification, thus there can be no genetically based racial differences. This argument overlooks the fact that racial classification can be based on small differences in gene frequencies at *multiple* polymorphic loci. Looking at a single locus gives us almost no taxonomically useful information, but looking simultaneously at multiple loci allows us to classify people into ethnic groups with accuracy that can approach 100% (Edwards, 2003; Tal, 2012). David Reich, who runs a leading genetics lab at Harvard University, recently acknowledged that Lewontin's argument has been used by geneticists to *"deliberately"* conceal the possibility of significant genetically based population differences:

> I have deep sympathy for the concern that genetic discoveries about differences among populations may be misused to justify racism. But it is precisely because of this sympathy that I am worried that people who deny the possibility of substantial biological differences among populations across a range of traits are digging themselves into an indefensible position, one that will not survive the onslaught of science. In the last couple of decades, most population geneticists have sought to avoid contradicting the orthodoxy. When asked about the possibility of biological differences among human populations, we have tended to obfuscate, making mathematical statements in the spirit of Richard Lewontin about the average difference between individuals from within any one population being around six times greater than the average difference between populations....But this carefully worded formulation is deliberately masking the possibility of substantial average differences in biological traits across populations. (Reich, 2018, p. 254)

Reich addresses another common argument used to reject the possibility of substantial race differences: even if there is genetically based variation "affecting cognition or behavior," these differences must be small because "so little time has passed since the separation of populations." He finds this argument untenable in the light of recent findings in genetics:

> The average time separation between pairs of human populations since they diverged from common ancestral populations, which is up to around fifty thousand years for some pairs of non-African populations, and up to two hundred thousand years or more for some pairs of sub-Saharan African populations, is far from negligible on the time scale of human evolution. If selection on height and infant head circumference can occur within a couple of thousand years, it seems a bad bet to argue that there cannot be similar average differences in cognitive or behavioral traits. Even if we do not yet know what the differences are, we should prepare our science and our society to be able to deal with the reality of differences instead of sticking our heads in the sand and pretending that differences cannot be discovered. (Reich, 2018, p. 258)

Section 2 explains why, from a scientific perspective, it is so difficult to determine the cause of group differences in intelligence. Although the genetic origin of race differences can only be definitively established when the fields of genetics and neuroscience are unified in a way that will not happen for several years, less definitive evidence that is currently available seems to implicate genes. Section 3 raises a question of values: should science be absolutely committed to the truth, or,

as some philosophers and scientists believe, are topics such as race differences in intelligence just too dangerous to face openly? Section 4 argues that utilitarian considerations favor open inquiry into the cause of race differences. Section 5 argues that truth has value independent of its consequences, though this value should be balanced against the potential dangers of truth. Section 6 argues that philosophers should help devise an ethical way to respond to group differences in intelligence in anticipation of the possibility that they will be shown to exist.

2. The scientific controversy

The mere fact that populations differ in IQ tells us little or nothing about the *source* of the difference. The difference could be due to genes, but it could also be due to some environmental factor(s)—education, nutrition, cultural bias in IQ tests, and so on. How can we disentangle the effects of genes and environment? Behavioral geneticists have two principal methods for probing the role of genes in phenotypic differences: (a) behavioral genetic studies on adoptees and twins, and (b) genome-wide association studies.

The heritability of a trait—the proportion of the variance in a population explained by genetic differences—can be determined by adoption and twin studies. If adopted children grow up to resemble their *biological* rather than their *adoptive* parents, this implicates genes. If (genetically identical) monozygotic twins raised together grow up to resemble each other more than (genetically related but nonidentical) dizygotic twins raised together, this, too, implicates genes (on certain testable assumptions, such as equal environments or, in the case of adoption studies, random placement into adoptive environments). Adoption and twin studies show that, in the contemporary United States, the heritability of IQ in adulthood is around 60 to 80% (Plomin & Deary, 2015).

Even if IQ has high heritability within racial groups, this does not imply that race differences are genetic. We cannot infer *between*-group heritability from *within*-group heritability. Lewontin (1970, pp. 7–8) illustrates this point with the following example. Suppose we have a sack of genetically variable corn seeds, and we plant a handful on two plots of soil. Conditions within each plot are carefully controlled so that each plant receives the same nutrients, sunlight, and so on. Since there is no variation in the environmental factors that affect growth, differences in height within each plot will be due entirely to genes—height will be 100% heritable. But suppose one plot contains a low concentration of nitrates. The corn in the low-nitrates plot will grow shorter on average for purely environmental reasons. Within-group (i.e., within-plot) heritability will be 100%, while between-group heritability will be 0%.

Heritability studies cannot show definitively that race differences in intelligence have a genetic cause. It is always possible that there is some hidden environmental factor(s)—a so-called "X factor," analogous to nitrates in the corn example—that explains differences between but not within races. (An X factor that explains the black–white IQ gap would have to affect all Blacks equally in order to preserve high heritability and shift the population mean downwards without changing the variance.) Nevertheless, the high within-group heritability of IQ can be part of a *package* of evidence for between-group heritability. This point was made by Jensen (1969), who was widely misunderstood as naively inferring between- from within-group heritability (see Sesardic, 2005, pp. 128–138). Expounding on Jensen's argument, Flynn (1980) writes that "the probability of a genetic hypothesis will be much enhanced if, in addition to evidencing high [heritability] estimates, we find we can falsify literally every plausible environmental hypothesis [i.e., X factor] one by one" (p. 40; quoted in Sesardic, 2005, p. 136). The obvious candidate X factor that could explain race differences is, of

course, *racism*. If racism lowers IQ, this could explain why the mean is shifted downwards for victimized groups. But, as Flynn argues, attributing differences to racism

> is simply an escape from hard thinking and hard research. 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; quoted in Sesardic, 2005, pp. 141–142)

Now, after decades of intensive searching, the X factor remains elusive. The adult black–white IQ gap has remained stubbornly constant at approximately one standard deviation (15 IQ points) among cohorts born since around 1970 (Murray, 2007). Dickens and Flynn (2006) report that "Blacks gained 4 to 7 IQ points on non-Hispanic Whites between 1972 and 2002" (p. 913), but these gains appear to be among Blacks born before the early seventies. Dickens and Flynn (2006, Figure 3) indicate that, in 2002, the black–white IQ gap among 20-year-olds was approximately one standard deviation, or 15 points. Nisbett (2017) writes that "Dickens and Flynn found [the Black–White gap in IQ to be] around 9.5 points," but this is only the gap if we include *children* (as Nisbett confirmed in a personal communication). More recent evidence indicates that the gap has persisted or even widened. Frisby and Beaujean (2015, Table 8) find a black–white IQ gap of 1.16 standard deviations among a population-representative sample of adults used to norm the Wechsler Adult Intelligence Scale IV in 2007. Intensive interventions can raise IQ substantially during childhood, when the heritability of IQ is low. But despite some misleading claims about the success of early intervention programs, gains tend to dissolve by late adolescence or early adulthood (Baumeister & Bacharach, 2000; Lipsey, Farran, & Durkin, 2018; Protzko, 2015). Adoption by white families—one of the most extreme interventions possible—has virtually no effect on the IQ of black adoptees by adulthood. Black children adopted by middle- and upper-middle-class white families in Minnesota obtained IQ scores at age 17 that were roughly identical to the African American average. Adoptees with one black biological parent obtained IQ scores that were intermediate between the black and white means (Loehlin, 2000, Table 9.3).[3]

To reiterate, the high within-group heritability of IQ combined with the failure to find an environmental X factor to explain the IQ gap does not show decisively that race differences are genetic, because it is possible that an X factor will be discovered in the future. However, the environmentalist theory of race differences has not, by normal scientific standards, been an especially progressive research program (in the sense of Lakatos, 1970). Environmentalists never predicted that the black–white IQ gap would, after reaching one standard deviation, remain impervious to early education, adoption, massive improvements in the socioeconomic status of Blacks, and the (apparent) waning of overt racism and discrimination. Commenting 45 years ago on

[3] The researchers who carried out the Minnesota Transracial Adoption Study argue that the "results...provide little or no conclusive evidence for genetic influences underlying racial differences in intelligence" (Waldman, Weinberg, & Scarr, 1994, p. 29). They give two main reasons for this conclusion. First, the mean IQ of Asian/Indian adoptees was lower than that of white adoptees, despite the fact that mean (East) Asian IQ is higher than that of Whites. This contradicts the hereditarian prediction that the rank ordering of the adoptees would reflect the rank ordering of their respective racial groups. Second, black children had worse early adoptive experiences than white children: Whites were placed with an adoptive family sooner and spent more time with them than Blacks (although half-black children were placed sooner than white children). Controlling for early experience explained a small proportion of the variance in IQ at age 17. But the fact remains that adoption did not have a significant effect on the IQ gap. Scarr (1998) later wrote: "The test performance of the Black/Black adoptees was not different from that of ordinary Black children reared by their own families in the same area of the country. My colleagues and I reported the data accurately and as fully as possible, and then tried to make the results palatable to environmentally committed colleagues. In retrospect, this was a mistake" (p. 230).

environmentalist theories that appeal to an X factor, Urbach (1974) noted that "any data in the world can be made consistent with any theory by invoking nameless and untested factors" (p. 134).

Let us turn to the second method for investigating the role of genes in development: genome-wide association studies (GWAS). Unlike heritability studies, GWAS can uncover *specific genetic variants*—or single-nucleotide polymorphisms (SNPs)—associated with IQ. In just the last couple years, GWAS has identified hundreds of such SNPs (Davies et al., 2018; Savage et al., 2018; Sniekers et al., 2017), which together explain around 11% of the variance in IQ (Allegrini et al., 2019).

If we find that the SNPs implicated in IQ are differentially distributed across racial groups, this would not necessarily imply that race differences in intelligence are genetic. SNPs might have different effects across races and environments due to gene–gene and gene–environment interactions. SNPs with no causal relation to intelligence can be genetically linked to SNPs that do have a causal relation in some populations but not others, so SNP–intelligence correlations may not always hold across races (Rosenberg, Edge, Pritchard, & Feldman, 2019). But if we find that many of the same SNPs predict intelligence *in different racial groups*, a risky prediction made by the hereditarian hypothesis will have passed a crucial test. Even then, however, GWAS will only establish a *correlation* between SNPs and IQ without revealing the causal chain linking SNP to phenotype. It would still be theoretically possible that these SNPs lead to differences in intelligence as a consequence of environmental factors (e.g., parenting effects) that can be manipulated so as to eliminate race differences. But if work on the genetics and neuroscience of intelligence becomes sufficiently advanced, it may soon become possible to give a convincing causal account of how specific SNPs affect brain structures that underlie intelligence (Haier, 2017). If we can give a biological account of how genes with different distributions lead to race differences, this would essentially constitute proof of hereditarianism. As of now, there is nothing that would indicate that it is particularly unlikely that race differences will turn out to have a substantial genetic component. If this possibility cannot be ruled out scientifically, we must face the ethical question of whether we *ought* to pursue the truth, whatever it may be.

3. Should science be absolutely committed to truth?

Truth was traditionally thought of—and is still often advertised—as the fundamental commitment of science (e.g., Dawkins, 2006). Some of the great heroes of science—most famously Galileo—are celebrated for resisting moral and religious authorities who feared that free inquiry would bring disaster. But the fact that the censors of previous eras were wrong does not mean that those of today cannot be right (Dummett, 1981, pp. 287–288). Sesardic (1992) (who ultimately defends free inquiry) observes that since the days when "truth was...regarded as the categorical imperative for scientists," we have "learned from some dramatic instances that there is no guarantee that...new knowledge [will] serve the good of mankind" (p. 129). Just as physics can produce an atom bomb with the potential to destroy us physically, perhaps social science can produce an equivalent—a discovery that would undermine the social order and destroy us just the same.

Many scientists claim to be absolutely committed to truth while at the same time advocating or tolerating varying degrees of suppression of controversial work. In his famous article, "The Responsibility of Intellectuals," Chomsky (1967) makes the bald statement that "it is the responsibility of intellectuals to speak the truth and to expose lies." Elsewhere, however, he says that scientists should not pursue the truth about group differences in intelligence because this could have dangerous social consequences (Chomsky, 1976, pp. 294–295). He says that while people

"surely...differ in their biologically determined qualities," it is wrong to investigate an association between group membership and IQ because to do so is to indicate "that the answer to the question makes a difference; it does not, except to racists, sexists, and the like" (Chomsky, 1988, p. 164; quoted in Cofnas, 2016, p. 486). "To anyone not afflicted with these disorders, it is of zero interest whether the average value of IQ for some category of persons is such-and-such" (Chomsky, 1988, p. 164).

Cofnas (2016) documented several prominent scientists and philosophers opposing the study or dissemination of findings related to group differences in intelligence. Howard Gardner, for example, who is known for his theory of multiple intelligences (Gardner, 1983), writes: "I myself do not condone investigations of racial differences in intelligence, because I think that the results of these studies are likely to be incendiary" (Gardner, 2001, p. 8). He also says that we should reject the work of hereditarian IQ theorists such as Arthur Jensen (whom he calls "the bad guys") because even if they

> turn out to be more correct scientifically than I am, life is short, and we have to make choices about how we spend our time. And that's where I think the multiple intelligences way of thinking about things will continue to be useful even if the scientific evidence doesn't support it. (Gardner, 2009, 45:11)

The leading intelligence researcher Robert Sternberg (2005) argues that good science is characterized by "taste in the selection of problems to solve" (p. 295), and that it is in bad taste to investigate the genetic basis of race differences. Sam Harris—a neuroscientist famous for aggressively promoting a "reason and evidence"-based worldview—recently interviewed *Bell Curve* (Herrnstein & Murray, 1994) coauthor Charles Murray. Harris accepted all of Murray's claims about hereditarianism. Yet in the introduction to the interview he said:

> I do remain skeptical about the wisdom of looking for cross-cultural or interracial differences in things like intelligence. I'm not sure what it gets you apart from a lot of pain. So many of the topics I discussed in the podcast with Murray are not topics I would ordinarily think about, or recommend that you think about. (Harris, 2017)

Although he does not argue that the information should be suppressed, he does believe that facts about group differences conduce to "pain" and we might be better off not thinking about them.

A degree of censorship is already in operation when it comes to findings supporting hereditarianism about group differences. Mainstream media coverage of the race-and-IQ controversy almost always falsely claims that there is a consensus among the relevant experts that hereditarianism has been refuted. In fact, anonymous surveys reveal that most experts on intelligence believe that there is a genetic component to race differences (Rindermann, Becker, & Coyle, 2016, 2020, Figure 3; Snyderman & Rothman, 1987, 1988). Research supporting both hereditarian and environmentalist explanations of race differences is routinely published in major psychology journals, particularly in psychometrics journals like *Intelligence* and *Personality and Individual Differences*. But work supporting hereditarianism can be much more difficult to publish and disseminate, and research testing the possible genetic basis of race differences is rarely funded. As James Flynn (after whom the "Flynn effect" is named) notes: "If universities have their way, the necessary research [on race and intelligence] will never be done. They fund the most mundane research projects, but never seem to have funds to test for genetic differences" (Flynn, 2012, p. 36). Flynn (2018) says that "scholars at one of America's most distinguished universities...admitted [to him] that they had never approved a research grant that might clarify whether black[s] and white[s] had equivalent genes for IQ" (p. 128). When he suggested some ways to test if there is a genetic

component to the black–white IQ gap, "they evaded the issue" (Flynn, personal communication). And because the state of the science is often misrepresented to the public (hence the common belief that there is no evidence at all for hereditarianism), findings are subject to de facto censorship. In some academic fields such as sociology and history, virtually all mainstream scholars refuse to consider the implications of group differences in intelligence for the problems they address. The questions before us are: Should research on group differences be conducted? Should findings related to differences be publicly disseminated? And, finally, what role should philosophers—who often act as intermediaries between science and the humanities, and between science and society at large—play in supporting or opposing work on the genetics of group differences?

Arguments in favor of restricting scientific research generally, or research on group differences specifically, are almost always utilitarian: They contend that free inquiry can, in some cases, be expected to have an overall negative effect on human well-being (e.g., Block & Dworkin, 1974; Chomsky, 1976; Dummett, 1981; Kourany, 2016). I argue that utilitarian arguments in favor of restricting research on group differences in intelligence have generally given short shrift to the potential and *actual serious harms that have resulted from the current practice of stigmatizing and dismissing controversial work*, and they have also ignored the potential benefits that might follow from openly confronting evidence about group differences.

I suggest that there are also at least two compelling *non*utilitarian reasons to allow free inquiry (compelling, that is, to those who are not utilitarians). First, truth is (to some degree) valuable independent of its felicific consequences. Proscribing the study of group differences undermines the integrity of huge swaths of work in biology, psychology, economics, sociology, history, and moral and political philosophy. If we accept that truth is intrinsically valuable, this is a cause for concern. Second, if not all groups have identical distributions of potential, then it is unjust to assume that some people must be blamed for average differences in performance among groups.

4. Utilitarian reasons to favor free inquiry

In 2001, virologists at the State University of New York at Stony Brook created a synthetic version of the poliovirus using strands of DNA purchased by mail order. They obtained the RNA sequence of the virus from the Internet, where it is freely available for all to see. Other researchers have published what are essentially instructions for recreating a smorgasbord of viruses that could potentially kill tens of millions of people were they to be released in the modern world (Kourany, 2016, pp. 784–785). In response, prominent medical authorities have called for restrictions on creating deadly viruses or publishing information that could help terrorists to do so (National Research Council, 2004).

Kourany (2016) asks: "Is there any reason new policy constraints like these for the life sciences should not be put into effect for the social sciences—new policy constraints that include research guidelines for weighing societal harms of research against societal benefits...?" (p. 786). As to the suggestion that the potential harm from deadly viruses is greater than the harm from research on intelligence differences, she urges us not to minimize the latter. Intelligence research

> has already been shown to cause significant harm...to lots more people [than a virus]—to all the people whose self-esteem, self-efficacy, ambitions, and successes are lessened as a result of direct or indirect exposure to the research or aspects of the research...; to all the people whose self-esteem, self-efficacy, ambitions, and successes are lessened as a result of the treatment they receive from others who have been directly or indirectly exposed to the research or aspects of the research; and so on—in short, to all or most women and, in the United States at least, to most minority men and many men of color in other parts of the world....Cognitive group differences

> research, then, arguably does pose harms to society near—perhaps even exceeding—the harms posed by the recent synthetic genomics research. (p. 787)

The reality of all these harms cannot, however, be taken for granted. As for the situation in the U.S., it is a well-established finding that the average levels of self-esteem in ethnic groups has, contra common wisdom, a *negative* relationship with average IQ. At least since the 1980s, Blacks have scored significantly higher than Whites, who in turn have scored higher than Asians, on measures of self-esteem (Bachman, O'Malley, Freedman-Doan, Trzesniewski, & Donnellan, 2011; Twenge & Crocker, 2002). Is the success of members of some ethnic groups impaired by hearing about research on intelligence differences? That depends on the strength of the effect of "stereotype threat." According to the theory of stereotype threat, telling people that their group has a low mean score on a cognitive test (i.e., priming them with a negative stereotype) impairs their performance (Steele & Aronson, 1995). In recent years, stereotype threat has run into trouble.

The "replication crisis" in psychology has revealed that many widely touted discoveries—particularly in social psychology and, within social psychology, *specifically* in regard to priming effects—may be completely false (Kahneman, 2012; Open Science Collaboration, 2015; Yong, 2012). Some psychologists have suggested that stereotype threat could end up being another casualty of the crisis. Ganley et al. (2013) failed to replicate the effect in three large studies of female mathematics performance. Using statistical techniques to test for publication bias, Flore and Wicherts (2015) found evidence that publication bias may "seriously distort the literature on the effects of stereotype threat" (p. 25) on math performance in girls. In a pre-registered study of more than two thousand Dutch high school students, Flore, Mulder, and Wicherts (2018) failed to find any evidence that gender stereotype threat affects math performance. There has not yet been an authoritative replication attempt of studies on the effect of priming with *racial* stereotypes. As of now, we certainly cannot say that stereotype threat "has already been shown to cause significant harm" (Kourany, 2016, p. 787). If stereotype threat is not real, then conducting research on group differences in IQ will not harm lower-scoring groups via a nonexistent effect. Even if it does turn out to be real, it is not clear that it is responsible for "significant harm," since the black–white IQ gap remains unchanged even when stereotype threat is eliminated. Stereotype threat is alleged only to *increase* the gap under controlled test-taking conditions (Sackett, Hardison, & Cullen, 2004; Steele & Aronson, 2004, p. 48). If stereotype threat turns out to be both real and to cause some devastating harm, that could raise ethical problems for intelligence research—but even its reality is far from being established.

There is also an important sense in which restrictions on publishing instructions to create deadly viruses and restrictions on intelligence research are not analogous. In the former case, the proposal is to *conceal* knowledge while being open about our motives: we should not (according to the proposals) publish instructions about how to make deadly viruses because we do not want terrorists to use them to make biological weapons. In the latter case, however, the proposal is to, if necessary, *misrepresent* the state of knowledge while being dishonest about our motives—to assert that there are no innate differences between groups while pretending that this is an established discovery of science. Given the disanalogy, we cannot so easily make an inference from the acceptability of limits on publishing work in virology to the acceptability of censoring or prohibiting work on intelligence, *even if* findings concerning intelligence differences could be shown to produce some sort of harm.

It is also easy to overlook the harms that have been caused by uncritical commitment to environmentalism. Kourany (2016) comments:

> Finding out that blacks have lower IQ scores than whites...could be the beginning of educational and training programs to work with the strengths and work on the weaknesses of every group to help make them the very best they can be, and even to use the special talents of each group to help the others. Finding these things out could be the beginning of innovative programs that support rather than undermine the right to equality. That this does not happen, or seldom happens, is a function of the...racism of society.... (pp. 783–784)

But the reason that these programs, which Kourany rightly says ought to exist, have never been created is not because of racism but *because of the taboo on talking about genetic differences among policy makers*. No mainstream politician can acknowledge that there are differences that might call for the creation of a program to "work with the strengths and work on the weaknesses of every [ethnic] group to help make them the very best they can be." It is *hereditarians* who have advocated these programs and *environmentalists* who have resisted them. The abstract to Arthur Jensen's 1969 paper—the first major modern defence of hereditarianism—says (written in the third person):

> Jensen examines other mental abilities [besides *g*] that might be capitalized on in an educational program, discussing recent findings on diverse patterns of mental abilities between ethnic groups and his own studies of associative learning abilities that are independent of social class. He concludes that educational attempts to boost IQ have been misdirected and that the educational process should focus on teaching much more specific skills. He argues that this will be accomplished most effectively if educational methods are developed which are based on other mental abilities besides IQ. (Jensen, 1969, p. 2)

Contemporary hereditarians have also called for tailored training programs (e.g., Gottfredson, 2005a; 2005b, p. 318; Lubinski & Humphreys, 1997). Accumulating evidence suggests that education can indeed be effective in raising specific cognitive skills (Ritchie, Bates, & Deary, 2015).

Decades ago, Jensen (1969) was vilified for predicting that Head Start—a program based on the environmentalist assumption that early intervention can permanently raise intelligence and academic performance—would not have its intended effects. In *The Bell Curve*, hereditarians[4] Herrnstein and Murray (1994) marshaled compelling evidence that Head Start was not working. In 2012, a congress-mandated report by the U.S. Department of Health and Human Services (HHS) stated that the effects of Head Start disappear by third grade (Puma et al., 2012, p. xxi)—that is, Head Start was not working. (Of course, the report did not endorse hereditarianism or recommend that Head Start be discontinued. It suggested that the program may confer some as-yet-undetectable benefits that last many years after participation.[5]) Nevertheless, in the years immediately following HHS's report, the government increased funding for Head Start by hundreds of millions of dollars (Enriquez, 2016). If we had followed Jensen's recommendation in 1969 to devote money to programs that were tailored to the strengths of different groups rather than to Head Start, around two hundred billion dollars would have gone to improving lives instead of accomplishing nothing that can be detected. This is one of many examples of how basing social policies on questionable scientific premises can have enormous opportunity costs.

4.1. Unknown dangers

It is impossible to know exactly what the consequences would be if race differences were proven to have a substantial genetic component and this finding became widely accepted. We ought to try to

[4] Herrnstein and Murray (1994) conclude that "it seems highly likely...that both genes and environment have something to do with racial differences" (p. 311). They do not give an estimate as to the relative importance of one or the other.
[5] Some scholars claim that early intervention programs, such as the Abecedarian Project, have produced lasting benefits. As noted, the evidence for these claims is often questionable (see Baumeister & Bacharach, 2000).

anticipate the potential harms and—keeping in mind the poor track record of even the most talented prognosticators of cultural trends—determine how seriously we should take them.

In the liberal West, many of our institutions, laws, and moral values are predicated on the assumption that race differences are either nonexistent or environmentally caused. If the hereditarian theory of race differences became widely accepted, it seems inevitable that *there would be significant cultural changes*, even if we cannot predict what those changes would be with any precision. Just about any radical change—even if it ultimately benefits the vast majority of people—will have *some* negative consequences. But the fact that change always has negative consequences is not *by itself* a compelling reason to preserve the status quo. Every instance of progress has negative byproducts: alternative energy puts coal miners out of work, teaching evolution causes distress to creationists, and so on. Many examples throughout history show that attempts to block scientific progress, for whatever reason, are generally both futile and ultimately harmful. Is research on race differences a special case? Would finding a genetic explanation for race differences pose a uniquely serious threat?

A common fear is that, if race differences were proven to have a genetic basis, this would cause people to turn to Nazism. Indeed, the study of race differences is often explicitly equated with Nazism. This fear seems to be based on a historical misunderstanding. Nazi ideology was not based on scientific discoveries. The Nazis were flagrant pseudoscientists whose research in biology and psychology was permeated with ideology. Contrary to a popular myth, both the Nazis and their ideological predecessors (such as Joseph Arthur Comte de Gobineau and Houston Stewart Chamberlain) rejected Darwinism. (Richards, 2013, magisterially documents the lack of a connection between Darwinism and Nazism.) Most important for the present discussion, Nazi scientists rejected mainstream intelligence testing and the concept of IQ (i.e., general intelligence, or the *g* factor derived by factor analysis), preferring a mystical view of intelligence tied to race. Erich Jaensch, an influential Nazi psychologist at the University of Marburg, claimed that IQ tests advanced the "supremacy of Bourgeoisie spirit" and would be a tool "of Jewry [to] fortify its hegemony" (Rindermann, 2018, p. 61). Jaensch's student Friedrich Becker advocated "intelligence measurement according to a national and typological point of view," and called for what was essentially a version of multiple-intelligences theory that accounted for "realism," "conscientiousness," "the character value of intelligence," and "practical intelligence" (Rindermann, 2018, p. 61).

The fact that the Nazis rejected Darwinism and the concept of IQ does not necessarily mean that the study of race differences could never inspire a dangerous political movement. It does mean, however, that we cannot point to Nazi Germany as an illustration of what is liable to happen when people adopt a biologically informed view of the evolution of intelligence. In Nazi Germany, "science" followed from ideology, not the other way around. Some contemporary neo-Nazis claim to find support for their Nazi-inspired views in Darwinism and IQ research, but their ideology tends to dictate their interpretation of the science—which is usually grossly uninformed.

There is a legitimate fear that findings on race differences could be widely misunderstood in ways that would promote racial discrimination, hatred, and conflict. Our tendency for "folk essentialism," which manifests in early childhood (Gelman, 2003), is a possible source of danger. People tend to explain why each entity has the properties it does by appealing to invisible essences shared by all members of a particular category that generate those properties (Gelman, 2003). Thus, children believe that rocks, men, women, and trees belong to categories whose members possess rock-ness, maleness, etc. The fear is that the discovery of race differences could reinforce our tendency to see members of different races as possessing different essences, which would in turn foster stereotypical thinking and discrimination.

There is reason to think that this dismal outcome is not inevitable. First, cross-cultural studies have found that "essentialist intuitions [vary] enormously across cultures and education levels" (Olivola & Machery, 2014, p. 499), and are thus highly plastic. Some evidence suggests, for example, that in cultures that recognize transsexual women as a third gender people do not essentialize males (Olivola & Machery, 2014). Perhaps we should not be too pessimistic about people's capacity to learn the error of essentializing race. Second, outside of certain university departments, virtually everyone believes that there are physical and psychological differences (on average) between men and women, and most people essentialize these differences. The fact that most people in the West essentialize the differences between men and women did not stop men from supporting women's rights, and did not prevent societies from making great progress toward that end. Even if people continued to (erroneously) essentialize races, this need not create an insuperable barrier to racial harmony.

We should also keep in mind that we have to weigh the danger of open inquiry *against the alternative*, which is opposing open inquiry. If there is a legitimate fear that disseminating findings about race differences could lead people to essentialize race and then lead to racial conflict, we have to consider the possibility that *concealing* these findings could lead to a similar or perhaps worse result. As long as people believe that race differences have a purely environmental cause, differences will, in practice, most likely be attributed to racism or institutional racism. Denying the possible genetic cause of race differences will not stop people from being focused on race. If people believe that members of certain races are victimized or benefited by racism, this could also foster essentialist thinking. It may be *more* likely to lead to racial conflict, since, as discussed below, some people will invariably be *blamed*, and whole races are likely to be held collectively accountable.

5. Nonutilitarian considerations in favor of free inquiry

5.1. The intrinsic value of truth

From a utilitarian perspective, there is nothing inherently wrong with falsehood. If not "productive of any material effects, [it] can never, upon the principle of utility, constitute any offense at all" (Bentham, 1823/1876, § 16.24). However, many utilitarians take it to be an *empirical* fact that the felicific consequences of falsehood tend to be negative, and they oppose it on those grounds. Singer (1996), for example, advocates freedom of inquiry as far as *what* can be investigated, and specifically opposes restrictions on intelligence research.[6] (Naturally, he opposes *methods* of inquiry that would harm human or animal subjects without being justified on utilitarian grounds.) Bentham (1823/1876) says that "combined with other circumstances, there is scarce any sort of pernicious effect which [falsehood] may not be instrumental in producing" (§ 16.24). When it comes to extolling truth, even philosophers who are not utilitarians often emphasize its instrumental usefulness rather than its intrinsic value. Clifford (1877/1999), citing practical dangers, famously suggests that it is immoral to believe anything without sufficient evidence. He reasons that even if no harm follows from a particular false belief, we must guard against cultivating a habit of credulousness, which would eventually lead to our destruction. When Frankfurt (2006) considers

[6] Although it is not entirely clear what his views are, Singer (2007) appears to believe that, given the state of the science (at least in 2007), advocating hereditarianism about group differences "raise[s] the suspicion of racism" because hereditarianism "lacks a solid scientific foundation." In this passage Singer was responding to James Watson, co-discoverer of the structure of DNA, who wrote that "the overwhelming desire of society today is to assume that equal powers of reason are a universal heritage of humanity. It may well be. But simply wanting this to be the case is not enough. This is not science. To question this is not to give in to racism" (Watson, 2007). Singer (2007) agrees that "questioning this assumption is not, in itself, racist."

why truth is important, the first thing that comes to his mind is that "truth often possesses very considerable practical utility. Any society that manages to be even minimally functional must have...a robust appreciation of the endlessly protean utility of truth" (p. 15). He expounds:

> Our success or failure in whatever we undertake, and therefore in life altogether, depends on whether we are guided by truth or whether we proceed in ignorance or on the basis of falsehood. [To] the extent that we recognize what dealing effectively with the problems of life entails, we cannot help loving truth. (pp. 35–36, 48)

All the reasons Frankfurt gives for why we should be concerned with truth connect directly or indirectly to its usefulness in helping us to satisfy our desires or otherwise achieve our goals.

There is also a long-standing tradition that truth has some intrinsic value, and that comprehending the truth and acting in conformity with it are worthwhile goals even if, in some cases, "he who increases knowledge increases sorrow" (Ecclesiastes 1:18). When Democritus said that he would "rather discover one causal law than be King of Persia" (Pearl, 2000, p. 41). he was probably expressing his deep commitment to truth rather than claiming that understanding leads to the greatest possible pleasure. Nozick (1974) took it for granted that most people (under normal circumstances) would reject a life of pleasure in an "experience machine" at the cost of being unaware that their perceived reality was a simulation and thus not really *true*.[7]

People do have different intuitions and views about whether truth is intrinsically valuable. But it seems likely that many professional philosophers and scientists are driven by a desire to attain some degree of true understanding of the world, ourselves, and our place in it. For these scholars, having wrong views is a tragedy even if they never suffer pain because of their mistake.

If there is meaningful, evolved, genetically based psychological diversity among geographic human populations, this has far-reaching implications for several areas of science (Winegard, Winegard, & Boutwell, 2017) as well as philosophy. We cannot understand the nature and evolution of intelligence unless we can explain why selection favored different levels of general intelligence, or specific intelligences, under different conditions. Since IQ is the best predictor of many individual and group outcomes that are studied in social science (Deary, 2012; Gottfredson, 2004; Haier, 2017; Jones, 2016; Rindermann, 2018; Warne, 2020), we will be in the dark about many sociological phenomena if we refuse to confront intelligence differences. Most theories in moral and political philosophy acknowledge the possibility of differences among individuals, but do not address the possibility of group differences. If holding ethical views that are consistent, justified, and applicable to the real world is valuable to us, then we must consider the implications of human diversity.

Those who believe the pursuit of truth to be valuable for its own sake recognize other, competing values. No sane person would pursue some interesting truth at the cost of destroying society or instigating genocide—the results feared by some opponents of research on race differences. The point is only that the intrinsic value of truth should at least be *part* of our moral calculation.

5.2. Justice

Linda Gottfredson (2005b) observes that "currently, racial parity in outcomes is often treated as the ultimate standard for fairness and lack of parity as a measure of White racism" (p. 317). Those who deny that there is evidence in favor of hereditarianism are forced to conclude that phenotypic differences between groups "must be artificial, manmade, manufactured. *Someone* must be at fault"

[7] "We learn that something matters to us in addition to experience by imagining an experience machine and then realizing that we would not use it" (Nozick, 1974, p. 44).

(p. 318). But if hereditarianism is true, then it may be that no one should necessarily be blamed for different average outcomes among groups. There is no theory of justice that says it is right to falsely blame a group of people for wrongs they did not commit because confronting the exculpatory evidence causes us discomfort.

Some environmentalists may blame "structural racism" for group differences in IQ.[8] Structural racism refers to the lingering effects of past racism that are embodied in our institutions, and which have pernicious effects regardless of whether or not living people are personally racist. Although the theory that IQ gaps are due to structural racism does not imply that living people are racist, it still implies that living people should be blamed in some ways for unequal outcomes. If the theory is correct, people who benefit from structural racism have a moral obligation to oppose it and counteract its effects. So long as racial gaps remain, people will be blamed for failing to extirpate structural racism.

6. Conclusion: The responsibility of philosophers

The strategy—advocated by some influential scholars—of stigmatizing, suppressing, or downplaying evidence in favor of hereditarianism about group differences has been tried and has not worked. Research on this topic has been done and the results are widely available. Major psychology journals continue to publish work that deals openly with group differences (though researchers still debate the relative contribution of genes and environment, and the question has not been settled definitively). Any measures that would be effective in preventing further work, such as those advocated by Kourany (2016), would have to be so severe that they would only attract even more attention to the findings they aimed to suppress. Science will carry on, and these questions will be answered. We should prepare *in advance* for the possibility that the genes underlying intelligence differences will not be distributed identically among ethnic groups. Failure to do this will only create a vacuum for "cranks rather than scientists" to opine on the nature and consequences of group differences (Anomaly, 2017, p. 293). Reich (2018) warns that if scientists "willfully abstain from laying out a rational framework for discussing human differences, [they] will leave a vacuum that will be filled by pseudoscience" (p. 258).

I have argued that the usual utilitarian reasons given for restricting intelligence research are not convincing and, in fact, there are strong reasons, both utilitarian and nonutilitarian, to favor free inquiry. For philosophers specifically, there is an additional consideration. For decades, the contribution of philosophers to this debate has consisted mostly in providing alternative explanations for evidence seeming to support hereditarianism about race differences (see Sesardic, 2000, 2005), and advocating various kinds of restriction and censorship (see Cofnas, 2016). This may be because hereditarianism is controversial, and philosophers are strongly disincentivized from pursuing lines of argument that lead to truly controversial conclusions. Testifying to how serious this problem is, Jeff McMahan, Francesca Minerva, and Peter Singer recently founded the *Journal of Controversial Ideas*, which will allow scholars to publish pseudonymously. Singer (2017) commented that "it's unfortunate that such a journal should ever be considered necessary to enable controversial ideas to be published, but perhaps we have got to the point where it is." It is not clear what kind of controversial ideas Singer had in mind, and the journal has not yet released its first issue. But it is hard to find a more controversial idea than hereditarianism about race differences in intelligence.

[8] It is not clear why structural racism would lead to higher mean IQs for East Asians and Jews. But whether the "structural racism" explanation for group differences is plausible is not the point at issue here.

There is a danger for the philosophical community in putting our credibility on the line over the claim that race differences are entirely environmental. If work on genetics and neuroscience within the next decade produces convincing evidence that differences in measured intelligence among groups have a significant genetic component, there will be no way to conceal this information. The hereditarian explanation will have to be accepted, and people will know that philosophers were on the wrong side of the issue both scientifically and morally: scientifically, because we are supposed to be careful, disinterested commentators on scientific controversies, not activists supporting only the politically popular side; morally, because we did not help lay the groundwork for responding in a moral way to these facts that we should have known might be coming.

Acknowledgments: Thanks to Timothy Bates, Rafael De Clercq, Hrishikesh Joshi, Neven Sesardić, Omri Tal, and two anonymous reviewers for *Philosophical Psychology* for helpful comments on earlier drafts of this paper.

References

Allegrini, A. G., Selzam, S., Rimfeld, K., von Stumm, S., Pingault, J. B., & Plomin, R. (2019). Genomic prediction of cognitive traits in childhood and adolescence. *Molecular Psychiatry, 24,* 819–827.

Anomaly, J. (2017). Race research and the ethics of belief. *Journal of Bioethical Inquiry, 14,* 287–297.

Bachman, J. G., O'Malley, P. M., Freedman-Doan, P., Trzesniewski, K. H., & Donnellan, M. B. (2011). Adolescent self-esteem: Differences by race/ethnicity, gender, and age. *Self and Identity, 10,* 445–473.

Baumeister, A. A., & Bacharach, V. R. (2000). Early generic educational intervention has no enduring effect on intelligence and does not prevent mental retardation: The infant health and development program. *Intelligence, 28,* 161–192.

Bentham, J. (1823/1876). *An introduction to the principles of morals and legislation* (New ed.). Oxford: Clarendon Press.

Block, N. J., & Dworkin, G. (1974). IQ, heritability and inequality, part 2. *Philosophy & Public Affairs, 4,* 40–99.

Ceci, S. J., & Williams, W. M. (2009). Should scientists study race and IQ? YES: The scientific truth must be pursued. *Nature, 457,* 788–789.

Chomsky, N. (1967, February 23). The responsibility of intellectuals. *The New York Review of Books*. Retrieved from http://www.nybooks.com/articles/1967/02/23/a-special-supplement-the-responsibility-of-intelle/

Chomsky, N. (1976). The fallacy of Richard Herrnstein's IQ. In N. J. Block & G. Dworkin (Eds.), *The IQ controversy: Critical readings* (pp. 285–298). New York: Pantheon Books.

Chomsky, N. (1988). *Language and problems of knowledge: The Managua lectures.* Cambridge, MA: MIT Press.

Clifford, W. K. (1877/1999). The ethics of belief. In *The ethics of belief and other essays* (pp. 70–96). New York: Prometheus Books.

Cofnas, N. (2016). Science is not always "self-correcting": Fact–value conflation and the study of intelligence. *Foundations of Science, 21,* 477–492.

Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D.,...Deary, I. J. (2018). Study of 300,486 individuals identifies 148 independent genetic loci influencing general cognitive function. *Nature Communications, 9,* 2098.

Dawkins, R. (2006). *The god delusion.* New York: Houghton Mifflin.

Deary, I. J. (2012). Intelligence. *Annual Review of Psychology, 63,* 453–482.

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

Dummett, M. (1981). Ought research to be unrestricted? *Grazer Philosophische Studien, 12/13,* 281–298.

Edwards, A. W. F. (2003). Human genetic diversity: Lewontin's fallacy. *BioEssays, 25,* 798–801.

Enriquez, B. (2016). FY 2016 Head Start funding increase. Retrieved from https://eclkc.ohs.acf.hhs.gov/sites/default/files/docs/policy-pi/2016-08/acf-pi-hs-16-03_0.pdf

Flore, P. C., Mulder, J., & Wicherts, J. M. (2018). The influence of gender stereotype threat on mathematics test scores of Dutch high school students: A registered report. *Comprehensive Results in Social Psychology, 3,* 140–174.

Flore, P. C., & Wicherts, J. M. (2015). Does stereotype threat influence performance of girls in stereotyped domains? A meta-analysis. *Journal of School Psychology, 53,* 25–44.

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

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

Flynn, J. R. (2018). Academic freedom and race: You ought not to believe what you think may be true. *Journal of Criminal Justice, 59,* 127–131.

Frankfurt, H. G. (2006). *On truth.* New York: Knopf.

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

Ganley, C. M., Mingle, L. A., Ryan, A. M., Ryan, K., Vasilyeva, M., & Perry, M. (2013). An examination of stereotype threat effect on girls' mathematics performance. *Developmental Psychology, 49,* 1886–1897.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

Gardner, H. (2001). The ethical responsibilities of professionals. Retrieved from http://thegoodproject.org/wp-content/uploads/2012/09/GoodWork2.pdf

Gardner, H. (2009). Intelligence: It's not just IQ. Rockefeller University, New York. Retrieved from http://www.youtube.com/watch?v=cP4CBpLNEyE&feature=related

Gelman, S. A. (2003). *The essential child: Origins of essentialism in everyday thought.* Oxford: Oxford University Press.

Gottfredson, L. S. (2004). Life, death, and intelligence. *Journal of Cognitive Education and Psychology, 4,* 23–46.

Gottfredson, L. S. (2005a). Suppressing intelligence research: Hurting those we intend to help. In R. H. Wright & N. A. Cummings (Eds.), *Destructive trends in mental health: The well-intentioned path to harm* (pp. 155–186). New York: Routledge.

Gottfredson, L. S. (2005b). What if the hereditarian hypothesis is true? *Psychology, Public Policy, and Law, 11,* 311–319.

Gottfredson, L. S. (2010). Lessons in academic freedom as lived experience. *Personality and Individual Differences, 49,* 272–280.

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

Harris, S. (2017). Waking up with Sam Harris #73 - Foridden knowledge (with Charles Murray). Retrieved from https://www.youtube.com/watch?v=Y1lEPQYQk8s

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

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