Towards a More Humane Genetics Education: Teaching about human genetic variation can reduce racial bias

Brian M. Donovan[1], Rob Semmens[2], Phillip Keck, Elizabeth Brimhall, K.C. Busch[3], Monica Weindling[1], Alex Duncan[1], Molly Stuhlsatz[1], Zoë Buck Bracey[1], Mark Bloom[1], Susan Kowalski[1]

[1]BSCS Science Learning
5415 Mark Dabling Boulevard
Colorado Springs, CO 80918

[2]Stanford Graduate School of Education
485 Lasuen Mall, Stanford, CA 94305

[3]North Carolina State University
College of Education, Box 7801
Raleigh, NC 27695

Main Text: 11631 words
References: 108
Figures: 4
Tables: 3
Supplemental: Yes

Correspondence: Correspondence concerning this article should be addressed to Brian M. Donovan, The Biological Sciences Curriculum Study, 5415 Mark Dabling Boulevard, Colorado Springs, CO 80918; E-mail: bdonovan@bscs.org

Author contributions: Donovan (PI) was responsible for the study concept, design of all RCTs, funding acquisition, development of all intervention materials, teaching the intervention in RCT 1, all statistical analyses, and manuscript writing. Semmens co-developed the study design for RCT 1 and the curricular framework for the first iteration of the intervention materials. Keck co-developed all intervention materials, taught the interventions in RCT 1, and assisted in the think-alouds and edits that produced the intervention materials for RCTs 2-3. Brimhall edited all intervention materials, taught the intervention in RCT 1, and implemented RCT 3 at her school site. Busch co-developed the first iteration of the intervention materials for the climate change control condition in RCT 1. Weindling and Duncan co-developed the intervention materials for RCTs 2-3, recruited teachers for RCTs 2-3, and aided in conducting RCTs 2-3. Stuhlsatz, Buck Bracey, Bloom, and Co-PI Kowalski assisted in the editing of intervention materials in RCT 3.

Funding: RCT 1 was supported through the Stanford Interdisciplinary Graduate Research Fellowship. RCTs 2-3 were supported by NSF Award # 1660985 - *Towards a More Human(e) Genetics Education: Exploring how Knowledge of Genetic Variation and Causation Affects Racial Bias among Adolescents*

Declaration of Conflicting Interests: The authors declare that they had no conflicts of interest with respect to their authorship or the publication of this article.

Abstract: When people are exposed to information that leads them to overestimate the actual amount of genetic difference between racial groups, it can augment their racial biases. However, there is apparently no research that explores if the reverse is possible. Does teaching adolescents scientifically accurate information about genetic variation within and between United States Census races reduce their racial biases? We randomized 8th and 9th grade students (N = 166) into separate classrooms to learn for an entire week either about the topics of: (i) human genetic variation; or (ii) climate variation. In a cross-over randomized trial with clustering, we demonstrate that when students learn about genetic variation within and between racial groups it significantly changes their perceptions of human genetic variation thereby causing a significant decrease in cognitive forms of prejudice. We then replicate these findings in two computer-based randomized controlled trials (RCTs), one with adults (N = 176) and another with biology students (N = 721, 9th-12th graders). These results suggest that teaching about variation in the domain of genetics has potentially powerful impacts on social cognition during adolescence.

Throughout history the science of genetics has been used to justify policies that helped one race by harming another (Jackson Jr. & Depew, 2017). Even today the belief that races are genetically dissimilar is used to justify inter-ethnic hostility (Kimel, Huesmann, Kunst, & Halperin, 2016), prejudice (Dar-Nimrod & Heine, 2011), segregative behavior (Williams & Eberhardt, 2008), and discriminatory policies (Soylu Yalcinkaya, Estrada-Villalta, & Adams, 2017). Evidence suggests that the biology curriculum plays a role in the perpetuation of this problem (Donovan, 2015b). For example, early 20th century biology textbooks taught students about the biological superiority of the white race (Donovan, 2015b; Morning, 2011; Willinsky, 1998). Although such arguments are absent from contemporary texts (Donovan, 2015b; Morning, 2011), randomized control trials (RCTs) have demonstrated a cause-effect relationship between the treatment of race in the biology curriculum and the development of racial biases (Donovan, 2014, 2016, 2017). When students learn about the prevalence of particular genetic diseases in specific racial groups during middle or high school biology classes it can unintentionally lead youth to perceive more genetic variation between races than actually exists (Donovan, 2017) and thus infer that racial groups differ in intelligence for genetic reasons (Donovan, 2014, 2016, 2017). In turn, this learning affects students' support for policies that redress racial inequality in education by influencing how students' explain racial disparities in education (Donovan, 2016, 2017).

A biology curriculum that perpetuates racial bias by unintentionally increasing inaccurate beliefs about racial difference is inhumane because it harms those who suffer from racial discrimination. We demonstrate that teaching scientifically accurate information about genetic variation within and between the US census races can reduce racial bias by undermining the belief that racial groups are discrete. These findings establish proof of concept for the hypothesis that human genetics education can be designed to create a more humane society by reducing the prevalence of racial bias. We therefore argue that human genetics education is not a socially neutral endeavor in a racialized society such as the United States (US).

The Humane Genetics Education Hypothesis

The purpose of a humane genetics education is to reduce the prevalence of racial bias by changing the way that students perceive human genetic variation. Such a hypothesis raises three important questions. First, how much biological variation actually exists between and within racially defined groups? Second, why would teaching students this genetic content lead to lower levels of racial bias? After all, the belief that races differ genetically is significantly associated with the belief that racial inequality is not worthy of redress because it is a natural and unchangeable product of genes (Brueckner, Morning, & Nelson, 2005; Donovan, 2015a; Jayaratne et al., 2006; Morning, 2011). Thus, it is reasonable to predict that teaching about human genetic difference will increase racial bias. Indeed, one might point to the long history of genetically-justified racial bias in education (e.g. Jensen, 1969) as a reason for why it is not wise to discuss genetic differences between races in school biology. Yet, history is filled with scientists who challenged racism by pointing out the genetic flaws in racialist beliefs (e.g. Beckwith, 2009; Feldman & Lewontin, 1975; Gould, 1996; Graves, 2015; Lewontin, 1972; Livingstone & Dobzhansky, 1962). Those scientists have argued that accurate understandings of genetic variation undermine the apparent validity of prejudiced beliefs (Jackson Jr. & Depew, 2017).

Prejudiced beliefs based in genetic thinking share a set of assumptions that are scientifically flawed (Jackson Jr. & Depew, 2017). First is the premise that people of the same race are genetically uniform. Second is the premise that people of disparate races are categorically

different. When these assumptions are combined with the belief that biologically influenced abilities are immutable, people will then argue that it is pointless to intervene socially to reduce racial inequality because race biologically determines ability. Knowing why these beliefs about genetics and race are flawed is the content of a humane genetics education.

The Content of a Humane Genetics Education

Any two humans share 99.9% of their DNA, which means that 0.1% of human DNA varies between individuals. Studies find that, on average, 4.3% of genetic variability in humans occurs between the continental populations commonly associated with US census racial groups (i.e. Africa, Asia, Pacific Islands, The Americas, Europe). In contrast, 95.7% of human genetic variation occurs between individuals within those same groups (Rosenberg et al., 2002; Rosenberg, 2011). This means that if we randomly pick two individuals from two different continental groups and compare them to two randomly picked individuals from the same group, we can expect that the former will be 4.3% more different from one another than the latter (Donovan, 2015a). These findings undermine the idea of intra-racial uniformity because they show that people of the same group are different in their variable DNA. They also show that racial groups are not discrete.

In fact, this same pattern is found repeatedly in studies of human variation (Boas, 1911; J. H. Relethford, 2002). There is more variability in skull shape, facial structure, and blood types within racially-defined populations (e.g. 89% for craniometric traits, 86% for blood types) than there is between them (e.g. 11% for craniometrics, 14% for blood types) (J. H. Relethford, 2002). Population genetic theory predicts that this same pattern will hold for any human trait that is not under different selection pressures in disparate populations (Edge & Rosenberg, 2015).

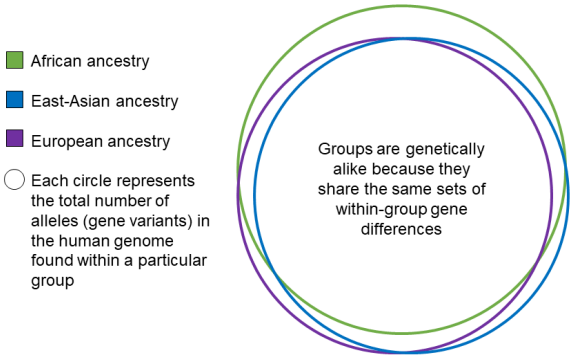
One known exception to this pattern is skin color, which varies more between populations than within populations (J. H. Relethford, 2002). Skin color changes continuously with distance from the equator (J. H. Relethford, 2002; 1997) because it is an adaptation to latitudinal differences in ultraviolet radiation (Jablonski & Chaplin, 2010). Therefore, there is no place on the planet where one can draw a line separating 'dark-skinned' populations from 'light-skinned' populations. Also, the gene variants associated with dark and light skin color are found in every major human population (Crawford et al., 2017). This means that the alleles associated with 'light skin' are found in 'dark-skinned' populations and vice versa (Crawford et al., 2017). The variants associated with dark or light skin are simply more prevalent in certain populations (Crawford et al., 2017). Both the continuous variation in skin color and the fact that groups share the genes associated with it undermine the idea that 'Blacks' and 'Whites' are discrete groups. Furthermore, because patterns of skin color variation do not match patterns of variation in other traits (i.e. blood types or skull shape), skin color cannot be used to make valid inferences about the geographic variation of other biological traits (J. H. Relethford, 2002). Thus, even if human behavior and cognition were genetically determined (and they are not, see Beckwith and Pierce (2018)), knowing a stranger's skin color would still not permit an accurate prediction about their abilities.

Since uniformity and discreteness beliefs about race are biologically flawed, it should come as no surprise that there has never been any agreement within the biological or anthropological sciences about whether human races are biologically real (Jackson Jr. & Depew, 2017; Morning, 2011). Even today there is no scientific consensus that race is biologically real (Foster, 2009; Kaplan & Winther, 2013; Kaplan & Winther, 2013; Weiss, 2008; Yudell et al., 2016). There is, however, a consensus that population genetic data does not support racialist claims (Coop, Eisen, Nielsen, Przeworski, & Rosenberg, 2014). For instance, physical anthropologists and population

geneticists have recently argued that genetic data: (i) refutes the notion that races are biologically real (Hunley, Cabana, & Long, 2016); (ii) cannot be used to support a biological theory of race without many additional philosophical arguments (Rosenberg & Edge, in press); and (iii) refutes the idea of racially pure populations (Reich, 2018). US racial categories are an artifact of our culture, history and personal beliefs, but not our biology (Markus & Moya, 2011).

This does not mean, however, that there are absolutely no biological differences between the socially-created racial groups used in the US. If this were true, then all variation would occur within populations and there would be no variation between geographic populations. In other words, *Homo sapiens* would be a single population. Instead, geographic populations of humans do differ. Yet, they do not differ in the ways that many people might think. Human groups differ in the proportion of people who have certain gene variants (Rosenberg, 2011). While some gene variants are unique to a single group (7.53%) (Rosenberg, 2011), on average none of those unique variants are possessed by more than 1.65% of any population (Rosenberg, 2011). Furthermore, on average, the amount of genetic difference between geographic groups of humans is about 7 times less than the genetic differences between populations of chimpanzees (Becquet, Patterson, Stone, Przeworski, & Reich, 2007) and about 14 times less than the genetic differences between populations of white-tailed deer (Graves, 2015). The relatively small genetic differentiation between human groups is a product of the finding that 92.47% of variants within the human genome are found in two or more continental groups of humans (Rosenberg, 2011). Thus, racial groups are genetically alike because they share similar sets of within group differences (Figure 1).

Figure 1. Genetic Differences Within and Between Groups



Note: This representation of human genetic variation was derived from data in Rosenberg (2011).

Nevertheless, the small proportional differences found when comparing the genomes of different US racial categories are often used to claim that racial inequality in the US is immutable because of genetics. Arthur Jensen (1969) was well known for making this kind of claim. In doing so, he overestimated the power of genes and the limits of scientific knowledge (Donovan, 2015a). For example, even when trait differences between individuals within a population are 100% associated with genes, and therefore entirely inherited, differences between populations can still be caused entirely by environmental factors (Feldman & Lewontin, 1975). If you estimate the heritability of skin color among white New Yorkers it will be high (Feldman & Lewontin, 1975). But, if you compare the skin color of white New Yorkers wintering in Florida to those who winter in New York, there will be a considerable difference in skin color between the groups that has no genetic basis (Feldman & Lewontin, 1975). To claim that racial disparities in education are caused by genetic differences between races on the basis of heritability statistics that are derived within

populations, as Jensen (1969) did, is a conceptual error (Feldman & Lewontin, 1975).

This error could also be an ideological distortion of the limits of scientific knowledge about behavior (Donovan, 2015a; Graves, 2015; R. Lewontin, 1996). Since the 1930s population geneticists and anthropologists have argued that cognitive and behavioral traits are malleable because genes produce different trait expressions in different environments because of gene-by-environment interactions (Boas, 1928; Dobzhansky, 1937). Modern research has shown that such traits are shaped through gene-by-environment interactions, epigenetics, environment, and chance (Caspi & Moffitt, 2006; Enoch, 2012; Graves, 2015; Haberstick et al., 2014; Keller, 2014; MacMahon, 1968; Okbay et al., 2016). Yet genome wide association studies have established that only 0.01% to 0.035% of variability in educational attainment is associated with any single gene variant (Okbay et al., 2016). We also know that scores on IQ tests change over time (Flynn, 1999) and their association with genetic factors depends heavily on environmental factors (Devlin, Daniels, & Roeder, 1997; Flynn, 1999; Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003). To claim that racial inequality in education is immutable because of genetics is no less of a distortion of scientific knowledge today than it was in the 1970s or 1930s.

Altogether, these findings demonstrate that it is incorrect to think that individuals of the same group are genetically uniform and that disparate groups are categorically different in their genes. Instead, a more biologically accurate view of human difference is that racially-defined populations are genetically alike in their variable DNA because each of these populations share similar sets of within group genetic differences (Figure 1). The purpose of a humane genetics education is to help learners make sense of these ideas in order to reduce their racial biases. Another purpose of a humane genetics education is to help learners understand that claims about race and genetics are not socially-neutral. When people claim that economic or educational inequalities between races are the result of genetic differences between races (e.g. Wade, 2014) they often misrepresent, willfully ignore, or misunderstand the complexities of human genetic variation (Coop et al., 2014; Donovan, 2015a; Feldman, 2014; R. Lewontin, 1996). A humane genetics education helps students understand that inequality is not the inevitable product of genes, but it is socially constructed, in part, by overly-simplified beliefs about genetics (R. Lewontin, 1996). Since prejudice is significantly associated with misunderstandings of human genetic variation (Donovan, 2017; Kimel et al., 2016; Williams & Eberhardt, 2008; Yzerbyt, Corneille, & Estrada, 2001) learning this genetic content for this humane purpose should reduce racial bias.

Reducing Racial Bias by Learning About Human Genetic Variation

The theory of change in a humane genetics education begins with how youth perceive human genetic variation. Studies estimate that 75% of college students taking introductory biology and genetic courses do not know that there is more genetic variation within ethnic groups than between them (Bowling et al., 2008). Among anthropology students, Hubbard (2017) found that 29% believed that there are more biological differences between two races than between individuals within a single race. A study by Donovan (2017) found that students (N = 135, 7^{th}-9^{th} graders) attending high socioeconomic status (SES) schools in the San Francisco Bay Area perceived 43% of genetic and phenotypic differences between racial groups and 57% within racial groups. It appears that students perceive far too much genetic variation across races and far too little within races. Due to the large mismatch between students' perceptions of human genetic variation and the actual scientific estimates of such variation, curriculum and instruction that

reduces the perception of variation between races should cause reductions in racial bias if such perceptions are causally implicated in the development of racial bias.

The relationship between perceptions of human genetic variation and stereotype endorsement is mediated by genetic essentialism (Dar-Nimrod & Heine, 2011; Donovan, 2015b). Genetic essentialism of race is a social cognitive bias which assumes that the genes *inherent* in people make same race individuals physically and behaviorally *uniform* and people of different races physically and behaviorally *discrete* (Andreychik & Gill, 2014). People who believe in the genetic *uniformity* of same race individuals have been found to believe that stereotypes apply to all group members (Yzerbyt et al., 2001). When people believe that racial groups are biologically *discrete* categories they also tend to endorse racial stereotypes (Bastian & Haslam, 2006) because discreteness beliefs facilitate category-based inductions about group members (Gelman, 2004). Finally, when people believe there are *inherent* differences in the genes of races, they attribute cognitive and behavioral differences between races to genetics (Donovan, 2016, 2017) because believing that groups cohere around inherent characteristics accentuates uniformity and discreteness beliefs (Yzerbyt et al., 2001). Therefore, if learning undermines either the belief that races are genetically discrete or the belief that same race individuals are genetically uniform, it could reduce stereotyping by impacting these essentialist beliefs.

To our knowledge, no randomized trials have tested the hypothesis that constructing an accurate understanding of human genetic variation can cause a reduction in these forms of racial bias. Previous studies have found that when biology education causes adolescents to perceive too much genetic variation between races it also causes an increase in genetic essentialism (Donovan, 2017). Thus, it should be possible to run this process in reverse and reduce racial bias. For instance, Aboud and Fenwick (1999) found that increasing adolescents' perceptions of observable differences (non-genetic) between individuals of the same race can reduce stereotyping among white students who are high in prejudice. Tawa (2016) found that exposing a single cohort of 31 adolescents to a five-hour intervention that taught about the biological similarity of people was associated with a reduction in genetic essentialism. Yet, since his study was non-experimental Tawa (2016) could not establish if his intervention caused these effects. Moreover, his study did not explore whether changes in genetic knowledge mediated the declines in genetic essentialism he observed. Hubbard (2017) found that undergraduates' (N = 296) misconceptions about genetic variation and race declined as they learned four key ideas about human genetic difference. But, like Tawa (2016), Hubbard (2017) could not attribute these declines to her intervention because she used a non-experimental pre-post design. Nor did she explore whether the declines in misconceptions were associated with changes in students' racial biases.

In sum, extant research and theory are consistent with the claim that genetic essentialism and stereotyping can be reduced by changing perceptions of human genetic variation. How, then, does one design and implement a humane genetics education?

Design and Implementation of a Humane Genetics Education

Science educators (e.g., Donovan, 2015b) have situated genetic essentialism within the theory of conceptual change (Strike & Posner, 1992) because it is a misconception that is negatively associated with biology learning (e.g. see Emmons & Kelemen, 2015; Evans et al., 2010; Opfer, Nehm, & Ha, 2012; Shtulman & Schulz, 2008). However, the cognitive work of changing genetic essentialist race conceptions through genetics education is socially situated for

students within schools. Thus, before attending to any theory of conceptual change, it is important to discuss why schooling is both an ideal and a difficult place to change racial beliefs.

Racial stereotyping develops, in part, because students use essentialist beliefs to make sense of racialized patterns observable within school (Bigler & Liben, 2007; K. Pauker, Ambady, & Apfelbaum, 2010; Kristin Pauker, Xu, Williams, & Biddle, n.d.). Refuting essentialist beliefs about race in school is thus a potentially powerful place to intervene in order to reduce prejudices in US society as people actively construct explanations for racial difference in school (Bigler & Liben, 2007; Carter, 2012; Donovan, 2014, 2016, 2017; Morning, 2011). At the same time, beliefs about inherent differences between groups may be difficult to change because adolescents use them to explain social disparities observable in their lives (Cimpian & Salomon, 2014).

Studies suggest that educational attempts to reduce racial prejudice in adolescent-aged students are more successful if they include information known to reduce prejudice that fits into the zone of proximal development of students (Aboud & Fenwick, 1999). If too complex, then using scientific information to correct biased thinking can backfire (Lewandowsky, Ecker, Seifert, Schwarz, & Cook, 2012). For example, when interventions challenge a myth with a scientific fact, they create an association between the myth and the fact, which increases the risk that students will conflate myth and fact, thereby leading to greater belief in the myth at a later date (Lewandowsky et al., 2012). This backfiring process, which we call the "Lewandowsky effect", is more probable when interventions seek to replace a simple explanation for the world with a more complex one (Lewandowsky et al., 2012) as we attempt here. The challenge, then, is to select the right information for students to learn, and to scaffold that learning to prevent backfiring.

Scaffolds implemented to prevent backfiring. Studies grounded in conceptual change have found that belief in misconceptions can be reduced through a skillfully designed refutational curriculum that triggers a misconception, labels it as incorrect, refutes it with evidence, and provides an alternative way of understanding the phenomena originally explained by the misconception (Guzzetti, Snyder, Glass, & Gamas, 1993; Lewandowsky et al., 2012; Van den Broek, 2010). A well-designed refutational approach does not simply tell learners about how to think about the world differently after a discrepant event; rather it introduces them to a more useful way of understanding the world. For an alternative explanation to be useful it must be plausible to the learner, which means that it must be congruent with, and connected to, other concepts the learner uses to make sense of the world (Lewandowsky et al., 2012). A good alternative explanation also helps the learner understand why prior explanations based in misinformation are flawed and why people tend to believe misinformed explanations (Lewandowsky et al., 2012). When these criteria are met there is a lower probability of backfiring.

The intervention created for this study was aligned with this specific refutational approach. It was designed to elicit biological essentialist beliefs about race (i.e. uniformity, discreteness) and then label them as biologically inaccurate through a discrepant event. Then, we prepared learners for an alternative explanation of the discrepant event by using contrasting cases (Schwartz & Bransford, 1998) of genetic data to help students understand why essentialism is genetically flawed and how anti-essentialist models of human difference are genetically accurate (see Table 1). Contrasting cases were used to scaffold this portion of the intervention because the use of anomalous data to create conceptual change can be ineffective when learners construe data through their pre-conceptions (Chinn & Malhotra, 2002). After learning from contrasting cases of genetic data, students were told an alternative explanation of racial difference. Then, we used scientific argumentation scaffolds (Osborne, Donovan, Henderson, MacPherson, & Wild, 2016) to support

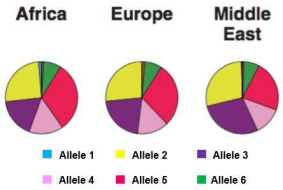
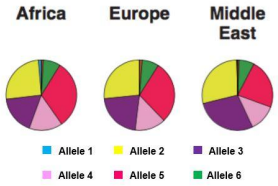
learners in an activity that asked them to critique an essentialist claim about race using their newly developed knowledge of human genetic variation. This last step of the learning process further supported conceptual change because argumentative critique has been found to support student sense-making in science (Ford, 2012).

The success of this intervention strategy depends, in part, on whether learners will adopt the alternative explanation about racial difference that we offer to them. We offered learners the idea that racial inequality is not a consequence of genetics, rather it is socially-constructed through incorrect genetic beliefs. Students were told that Americans develop incorrect ideas about human genetic variation for two *possible* reasons. First, there is a lot of misinformation about race and genetics in our culture. This misinformation has been promulgated by individuals who have a racially-biased social agenda. Such individuals try to convince other people that racial stereotypes are caused by genetics because they want to mislead people into thinking that racial inequality is natural and therefore not worthy of redress. Next, we told students that one possible reason why people tend to believe misleading information about race is because people live in segregated communities in the United States. Since segregation causes social isolation, it prevents people from seeing that individuals in another group are similar to them. It also prevents them from seeing that they differ from people in their own group in the same way that people in another group differ from each other. If individuals never get to have these social interactions, because of segregation, then it will be difficult for individuals to see that people of different races are alike because they share similar sets of within group differences. Therefore, people will tend to believe misinformation about race and genetics because no one has ever taught them an alternative way of thinking about genetic variation within and between races (e.g. the image in Figure 1).

The importance of contrasting cases. Adoption of this alternative explanation is more probable if students possess prior knowledge for understanding it (Alvermann, Smith, & Readence, 1985). Yet learners do not possess well-differentiated prior knowledge about genetic variation that would allow them to understand why essentialism is flawed (Bowling et al., 2008; Donovan, 2017; Hubbard, 2017b). We therefore designed our intervention using Schwartz and Bransford's (1998) "time for telling" framework to help learners differentiate genetics information that is consistent with, or which conflicts with, genetic essentialism. Schwartz and Bransford (1998) argue that opportunities to analyze information in sets of contrasting cases helps learners to perceive features that make a case distinct, thereby helping the learner differentiate information so it can be used in subsequent learning activities, such as reading informational texts (Schwartz & Bransford, 1998). We used contrasting cases to help learners see how genetic data is inconsistent with genetic essentialism so that students could comprehend the complex ideas that were presented in our alternative explanation for racial difference. Table 1 shows an example of this approach.

The contrasting cases introduced learners to counterfactual sets of data that either supported or refuted claims linked to belief in biological essentialism of race. For example, when learners were evaluating a claim about within group variation, such as "same race individuals are genetically different", they would first be presented with a representation of genetic data that would perfectly support the claim that same race people are genetically different (Table 1a). Then, they would be presented with data in the same graphical format that would perfectly refute this claim (Table 1b). These counterfactuals helped learners to differentiate patterns of human genetic variation by aiding their perception of how data would have to look to be consistent or inconsistent with an essentialist model of human difference. Learners were told that these contrasting cases were fictional and were created to help them make sense of the actual genetic data (Table 1c).

Table 1. Time for Telling Data Interpretation and Argumentation Scaffold

A. Counterfactual 1	 **Within group text:** If human genetic data looked like this, then it would support the claim that individuals of the same race are genetically different**.** In Figure 1, we see pie charts for three different ancestral groups. Within each ancestral group about 20% of individuals have each of the five different alleles. Because each ancestral group has five alleles each group has genetic variety. **Between group text**: Because each group has the same sets of alleles in the same proportions, the groups are genetically identical.
B. Counterfactual 2	 **Within group text:** If human genetic data looked like this, then it would refute the claim that individuals of the same race are genetically different. In Figure 2, we again see pie charts for the same three ancestral groups. In each ancestral group, every individual has the same allele. Since everyone from the same ancestral group would have the same allele, there would be no genetic variety within the ancestral group. **Between group text:** Since each group has entirely different alleles, the groups are categorically different. They share no genes at this locus.
C. Argumentation	 **Within group text:** Now let's look at real data from a study on human genetic variation (data from Rosenberg, 2011). Each pie chart shows the different alleles found at the same location on a chromosome within an ancestral group. This pie chart tells us that individuals of African ancestry might have any one of six different alleles at the same location in their DNA (represented by six different colors). The size of each pie slice shows the percent of individuals that have an allele. 33% of individuals with African ancestry have the red allele, 25% have the yellow allele, and about 5% have the blue allele. The same genetic variability can be seen in people of European ancestry and in people of Middle-Eastern ancestry. **Does the real data support or refute the claim that same race individuals are genetically different?** **Between group text:** Last time we looked at the number of pie slices in each chart, but this time we want to look at the size of each pie slice. The size of each pie slice is the percentage of individuals in each ancestral group that has a given allele. If you look at all 3 groups, you can see that they all have the dark purple allele. But, each group differs in the percent of individuals who have the dark purple allele. About 20% of individuals with African or European ancestry have the dark purple allele. But, 33% of individuals of Middle-Eastern descent have it. **Does the real data support or refute the claim that different races are genetically alike?**
D. Time for Telling	 **Within group text - What does this data tell us?** We can see that for each ancestral group, individuals in the group could have any one of several different alleles at a given location on a chromosome. The same pattern is found elsewhere in the human genome. Therefore, scientists have concluded that individuals of the same race are genetically different from each other. **Between group text - What does this data tell us?** We can see that individuals from different ancestral groups have the same set of alleles at a given location on their chromosomes. But, the percent of individuals in each pie chart who have a specific allele varies from group to group. This means that if you place one pie chart over another they match up very well in color but less well in the size of each pie slice. Thus, different races are genetically alike because they share similar sets of alleles. In fact, 92.47% of alleles in the variable part of the human genome are shared by two or more groups.

Learners were then given the actual genomic data in the same graphical format along with a text that explained how to interpret it (Table 1c). The real data resembled the anti-essentialist counterfactual, but it did not perfectly match it. Instead, the real data shared a deep structure with the anti-essentialist counterfactual. Students were thus instructed to argue whether the real data more closely matched the essentialist or anti-essentialist data. Then, learners were asked to use the real data to make an argument either for or against the claim that "individuals of the same race are genetically different" (Table 1c.). After, learners were told how to interpret the data (Table 1d.).

Using analogical induction to help learners identify the deep structure shared between the anti-essentialist counterfactual and real genetic data should result in more efficient learning of the anti-essentialist model (see Shemwell, Chase, & Schwartz, 2015). It should also increase the probability that learners transfer this model to other portions of the intervention (see Shemwell, Chase, & Schwartz, 2015). For instance, when exploring between group variation we asked learners to attend to proportional distinctions across pie charts in Table 1. This provided learners with an opportunity to see that groups share the same sets of gene variants, but each group is unique because it contains different proportions of people who possess each variant. This sets up a time to tell learners about a new deep structure in the data, which is the idea that different races are genetically alike because they share similar sets of within group gene differences (Figure 1). After helping learners make sense of these ideas, they were told the alternative explanation described above and then asked to engage in the critique portion of the intervention described earlier.

Summary. In essence, we use contrasting cases to scaffold data interpretation and scientific argumentation to create a "time for telling" students about an alternative, non-essentialist, explanation of human genetic difference. Our framework uses the principles of refutational texts, analogical transfer, and argumentative critique in order to facilitate conceptual change and reduce the probability of backfiring. At present, no studies have demonstrated that learning information about the complexities of human genetic variation will actually reduce racial bias. Yet, as we have argued, there are good reasons to think that learning about variation in the domain of genetics can powerfully influence social cognition. At the same time, the cognitive complexities of a humane genetics education lend themselves to misinterpretation in US culture, where ideas about genetic variation and race are usually used to rationalize inequality rather than challenge it (Morning, 2011). Since those rationalizations are constructed and practiced in schools (Bigler & Liben, 2007; Donovan, 2017; Morning, 2011; Willinsky, 1998) it is a worthy cause to explore if racial beliefs can be changed through a more humane genetics education.

Research Questions and Hypotheses

The main research question driving our study is thus: Does learning scientifically accurate information about genetic variation within and between US census racial groups cause a significant reduction in genetic essentialism and racial stereotyping? Our first two analyses test the predictions that teaching students scientifically accurate information about human genetic variation causes students to: (H_1) perceive proportionally less genetic variation between US census races relative to the total variation in humans (i.e. within and between races); and (H_2) exhibit less racial bias (i.e. endorsement of biological essentialism and racial stereotypes). Then, we directly test the mediational hypothesis that a humane genetics education reduces racial bias by changing perceptions of human genetic variation (H_3). Afterwards, we test whether reductions in racial bias are reversed over time because of the backfiring effect described by Lewandowsky et al. (2012).

Methods

To test our hypotheses, we use three different RCTs. In RCT 1, we estimate the impact of learning about human genetic variation over an entire week using a group-based version of our intervention. Then, in RCTs 2-3 we use a 45-minute computerized version of our intervention to explore the effects of a humane genetics education in a sample of adults and in a sample of

adolescents from a geographically diverse set of schools. We meta-analyze the effect sizes from these RCTs to explore their reproducibility. Then, we test whether a single mediating mechanism can account for the findings in each RCT. Finally, we test for 'Lewandowsky effects' by exploring the duration of treatment effects over three weeks using a subset of our third sample (n = 283).

Sample Descriptions of RCTs 1-3

RCT 1. We recruited students (N = 166) from a high SES private middle school (n = 52, 8th graders) and a high SES public high school (n = 114, 9th graders) in the San Francisco Bay Area (M_{age} = 14.3, SD_{age} = 0.74, % Female = 53%). A minority of students (6.8%) at the high school were eligible for free and reduced-price lunch (FRPL). The sample self-identified as White (48%), Mixed-race (19.88%), Asian (18.1%), and Hispanic (5.42%), with fewer students identifying as Black and Pacific-Islander (1.2% each). The remaining students declined to pick one of the US census categories (6.02%). Students were not paid for their efforts.

RCT 2. The purpose of RCT 2 was to pilot the computerized version of our intervention to test if it worked as intended before using it with students. We recruited adults (N = 176) from Amazon's Mechanical Turk website. Participants were paid $9.00 for their efforts. The sample of adults self-identified as White (76.7%), Black (13.6%), Mixed-Race (3.9%), Hispanic (3.9%) and Asian (1.7%). The majority of participants self-identified as female (63%) and the average age was 41 years old (SD = 12.24, range 21-73). Participants identified their political orientation as Republican (26%), Independent (24.43%), and Democrat (49.57%). Participants had either never finished high school, graduated high school but did not finish college, or earned associate degrees.

RCT 3. This sample of 9th-12th graders (N=721) was recruited from five schools. Two were public high schools in major cities in Colorado (33.34%). In California, the one public high school was located in the San Francisco Bay Area (41.19% of sample; same school in RCT1). We also sampled one public high school in the Greater Boston, Massachusetts area (20.39% of sample) and one private high school in the Washington DC area (4.99% of sample). FRPL in the two Colorado high schools was high at one site (FRPL = 66%) and low at the other (FRPL = 12.1%). FRPL for the remaining schools was low (Washington DC = 0%, Boston = 25.9%, California = 6.8%). The percent of white students at each school ranged from 71% in Boston to 20% at one of the Colorado sites. Participating students self-identified their race as White (61.7%), Asian (19.8%), Mixed-Race (9.9%), Hispanic (4.9%), Black (2.4%), Pacific-Islander (.55%), and American-Indian (.4%). The mean age of students was 15 (SD = 1.02, range 14-18) and 47.7% identified as female. Students were in 9th grade (54.1%), 10th grade (19.6%), 11th grade (25.9%), and 12th grade (.4%).

We attempted to recruit twice as many schools, especially those serving lower income students, more racially diverse populations, or those located in more politically conservative areas. However, district research offices for these schools rejected our study. One district research coordinator even rejected the study in an email because its intended purpose was to assess "psychological bias based on racial stereotyping". While this sample is geographically diverse, it is predominantly high SES, majority white, and it includes schools that had district research priorities that were aligned with our research goals. Students were not paid for their efforts.

Treatments Used in RCTs 1-3

RCT 1. A full description of the race and human genetic variation intervention used in RCT 1 can be found in the supplemental. In brief, the RCT 1 intervention taught students about the amount of genetic and phenotypic variability within US census races and between them. It also taught students how people tend to misrepresent those differences when claiming that race is biologically real. In contrast, the climate variation control condition taught students about the amount of temperature and precipitation variability within and between climate zones. Likewise, it taught students how people tend to misrepresent those differences when claiming that climate change is not real. Each intervention was designed to align with the core ideas, cross-cutting concepts, and practices of the Next Generation Science Standards (NGSS, 2016).

Table 2 provides an overview of how the treatments differed in RCT 1. Both treatments employed the same instructional framework (outlined in our conceptual framework) and both involved the same time spent on each task. The two interventions differed only in content objectives. We used a unit on climate variation for a control condition for two reasons. First, this unit taught about the concept of variation just like the race condition. Second, it is a politically controversial issue just like the topic of racial difference. This treatment-control contrast ensures that both groups learn about controversial topics in the media associated with scientific estimates of variation, thereby controlling for cognitive and political confounding. This is important because biological beliefs about race are ideological issues (R. Lewontin, 1996). To not control for ideology would undercut our claim to have run randomized controlled trials.

It might be argued that a more clinically-relevant control condition would involve business-as-usual (BAU) materials involving race and genetics, such as learning about the prevalence of genetic diseases in various racial groups. However, previous studies have found that such materials increase genetic essentialism (Donovan, 2014, 2016, 2017). Other studies have found that merely exposing individuals to genetic information primes belief in genetic essentialism (Lynch, Bevan, Achter, Harris, & Condit, 2008). Since we wanted to have an inert control we did not use either of these BAU controls. Arguably, the effects we estimate in this paper are smaller than those that will be found if our intervention is compared to such BAU controls in the future.

Table 2. Conceptual differences between treatments

	Core Ideas Taught	Day	Core Ideas	NGSS standards
Human Variation	1. Scientists do not agree about whether race is biologically real. 2. 99.9% of the DNA between any two humans is identical. When geneticists look at the variable portion of human DNA (0.1%) they find: (i) 95.7% of differences are between people of the same race (ii) 4.3% of differences are between people of different races 3. Skin color changes continuously as one moves away from the equator. But, there is more variation in skin color across races than within races. 4. When people make arguments about the superiority of one race over another they tend to overestimate the amount of genetic difference between races.	1. 2. 3. 4. 5.	1 2.i. 2.ii., 3 2.i, 2.ii 4	LS3A LS3B LS4A
Climate Variation	1. Scientists agree that the climate is changing. 2. Weather and the climate are different concepts. 3. When scientists support claims about climate change they use data on climate variation and not weather variation. For example, if we look over the last 100 years: (i) We can see that daily and monthly temperatures and precipitation change – this weather variation occurs within a climate zone and cannot be used to evaluate claims about climate change (ii) But, when we look across large land areas and periods of time greater than 30 years we see a continuous increase in the average temperature and precipitation across the United States – this climate variation data can be used to evaluate claims about climate change 4. When people make arguments that climate change is not real they tend to incorrectly use evidence about the weather to evaluate claims about climate.	1. 2. 3. 4. 5.	1, 2 2, 3 2, 3 4 4	ESS2D ESS3D

RCTs 2-3. For these studies we distilled the ideas in Table 2 into two 45-minute computer-based interventions using the learning approaches laid out in our conceptual framework (Table 1). The interventions were delivered through the Qualtrics platform. These computer-based interventions were vetted by the advisory board of our NSF grant, which included a population geneticist, sociologist, educational methodologist, developmental psychologist, two science education professors with expertise in functional and socio-linguistics, and a professor of gender and equity studies. Our materials were also vetted by three cooperating teachers and we piloted them with 8th graders and 12th graders by performing think-alouds. This process resulted in two interventions targeting the core ideas in Table 2 that were written at the same grade level (8th grade). Links to the actual computerized interventions can be found in the supplemental.

Dependent Variables in RCTs 1-3

Timing of Measurements in RCTs 1-3. In RCT 1 all instruments were administered at one-week intervals through a computer-based survey. We measured the students at baseline in RCT 1 in order to estimate the psychometric properties of our instruments and establish their convergent and discriminate validity. In RCTs 2-3, we only measured the two dependent variables after individuals were treated with the interventions in order to avoid Solomon effects (Solomon & Lessac, 1968). In RCT 3 at our California school site, we were able to collect an additional delayed post measurement of perceptions of human genetic variation and racial bias 3 weeks after students were treated in order to explore intervention backfiring via 'Lewandowsky effects'.

Perceptions of human genetic variation (H_1). We used the perceptions of biological variation measure (RCT 1: $\alpha = 0.90$; RCT 2: $\alpha = 0.88$; RCT 3: $\alpha = 0.86$) which was developed and validated to measure adolescent perceptions of genetic variation within and between races (Donovan, 2017). In the first four items of this measure, students are presented with a ten by ten matrix of purple circles that represent 100 people of the same race (labeled as either "White," "Black," "Asian," or "American-Indian"). The dots in these diagrams changed color randomly to become more heterogeneously colored as students moved a slider bar under the picture on a scale of 0-100% (More color variation = more within group variation = higher score). Students were asked to move the bar to a location to represent the percentage of DNA that differed between individuals of the same race. Next, students were presented with a series of six overlapping Venn-diagrams, one set for each combination of the four US census races listed above, which could be moved over each other to represent the biological difference between two groups. Students were instructed to make the Venn diagram represent the percentage of genetic material shared between the two labeled groups. Less overlap of the circles equated with a higher percent difference.

A principal components analysis supplemented with a parallel analysis demonstrated that these two item types, between and within items, were statistically discriminable (see supplemental). As in previous studies (Donovan, 2017), we averaged the between race and within race questions separately. Then, we applied Equation 1 to these data, which yielded a single proportion for each student that could take on any value between 0 and 1.

Equation 1:

$$\frac{\text{Perceived differences between races}}{\text{Perceived differences between races} + \text{Perceived differences within races}}$$

Equation 1 implies that students can increase in score on the instrument either by making the dot-diagrams less colorful and/or by placing the circles in the Venn diagrams further apart. Higher scores on this instrument (0-100%) indicate that a student perceives a greater proportion of differences between races relative to the total variation perceived within and between races. Prior to the start of the experiment, students in RCT 1 perceived, on average, 42% of human genetic variation between races (SD = 24%). Adults in the control condition of RCT 2 perceived 37% of human genetic variation between races (SD = 27%). Students in the control condition in RCT 3 perceived 36.5% of human genetic variation between races (SD = 28%).

Racial bias (H_2). To reduce Type I error we created a composite measure that combined a racial stereotyping instrument (see Levy, Stroessner, & Dweck, 1998) with items from two different genetic essentialism of race instruments (see Parrott et al., 2005; Williams & Eberhardt, 2008) to assess racial bias (as defined in Engberg, 2004). The instrument (RCT 1: $\alpha = 0.85$; RCT 2: $\alpha = 0.85$; RCT 3: $\alpha = 0.84$) assessed agreement with items such as: "Racial groups are primarily determined by biology"; "two Black people will always look more similar to each other than a Black person and a White person ever would"; "Racial differences in academic ability are caused by genetics"; "A randomly picked black person will be unintelligent". Items in each scale were standardized (Z-scored) and then averaged into a single variable of racial bias. Higher scores indicate greater bias. For example, in the supplemental, we show that students in RCT 1 scoring higher on this measure exhibited significantly greater levels of ingroup favoritism and intergroup anxiety. They were also more certain that they knew people who believed that racial inequality is natural, cannot change, and is not worthy of governmental redress. As expected, students scoring higher on this measure also perceived significantly more genetic variation between racial groups in RCT 1 ($r = 0.19$, $p = 0.017$) and in RCT 2 ($r = 0.26$, $p = 0.016$) and 3 ($r = 0.13$, $p = 0.017$).

Cross-Over Trial Design of RCT 1

Individual students were randomly assigned to learn about human genetic variation (the treatment) or climate variation (the control) for five, 45-minute, periods. We randomly assigned students within classrooms into two new classrooms so that students could learn in groups without introducing cluster level-correlations that reduce statistical power. Table 3 describes the design.

Table 3. Experimental design of RCT 1

School	Randomly Assigned Groups	Day 0	Intervention at Time 1 (T1)	Day 7	Intervention at Time 2 (T2)	Day 14	Day 21
Middle School	Race at T1 & Climate at T2 (R1C2)	Measure 1	Race (Teacher1)	Measure 2	Climate (Teacher1)	Measure 3	Measure 4
	Climate at T1 & Race at T2 (C1R2)		Climate (Teacher2)		Race (Teacher2)		
High School	Race at T1 & Climate at T2 (R1C2)		Race (Teacher3)		Climate (Teacher3)		
	Climate at T1 & Race at T2 (C1R2)		Climate (Teacher1)		Race (Teacher1)		

Formally, this design is called an individually randomized trial with clustering (IRTC) because individuals are randomized to experimental arms and then treated as a group (Kahan & Morris, 2013; M. J. Weiss, Lockwood, & McCaffrey, 2016). We conducted a cross-over IRTC

where half of the students were assigned to receive the human genetic variation and race intervention first, then the climate variation intervention second (R1C2). The other of half of students received the climate variation intervention first, then the race intervention second (C1R2).

After being randomized into treatment groups, students learned in separate classrooms for two weeks at each school. Random assignment created a total of 16 new classrooms. Because two different teachers implemented both interventions at each time point, we include fixed effects for teacher in our statistical modeling to control for time invariant differences between teachers.

Parallel Trial Design of RCTs 2 and 3

In our two computer-based RCTs, we block randomized individuals into experimental arms within self-identified racial groups. In RCT 3 there was an additional layer of block randomization because students were randomized to experimental arms within classrooms by self-identified race.

Analysis Summary

We report treatment effects on Z-scores of all variables, except in the mediation analyses. Standard errors are calculated at the person level because each RCT was person-randomized and there was a low intra-class correlation across classrooms for both dependent variables (e.g. prior to treatment in RCT 1, ICC = 0.02-0.071, in RCT 3, ICC = .009-.066). In the supplemental we further describe our statistical methods and demonstrate that random assignment in each RCT produced baseline equivalence. There we also show that there was no attrition and that missing data was missing at random. We also report findings from models using multiple imputation.

In RCT 1, where we test H_1 and H_2, we report intention to treat estimates from marginal regressions that modeled within subject correlations though an unstructured covariance matrix. To test the reproducibility of the findings supporting H_1 and H_2 observed in RCT 1 we compare them to the findings of RCTs 2 and 3 using a DerSimonian and Laird (1986) random effects meta-analysis. For our mediation hypothesis (H_3) we use a Sobel-Goodman test augmented with bootstrapping (N = 5000 replications) in each RCT. Finally, to explore backfiring we test if effects reverse after three weeks using a subset of our third sample and multivariate regression.

Results

Initial Support for H_1 and H_2 in RCT 1

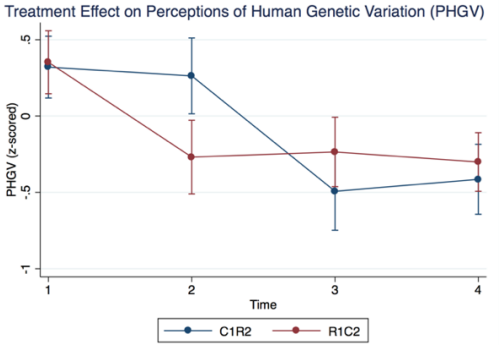
We predicted that the race intervention would cause declines in perceptions of human genetic variation and racial bias. Compared to the climate variation intervention (C1R2), the race variation intervention (R1C2) caused greater declines in the perception of genetic variation between races (χ^2 (1) = 7.10, p = 0.007) and in levels of racial bias (χ^2 (1) = 9.57, p = 0.002) (ps < .05, after Bonferroni adjustment) between time points one and two (Figure 2A-B). After treatment at time point 2, students in the race condition (R1C2) also perceived significantly less genetic variation between races (β = -.524, SE = .191, p = 0.006, d = -0.59) and also exhibited significantly less racial bias (β = -.308, SE = .128, p = 0.016, d = -0.48) (ps < .05, after Bonferroni adjustment) than students in the climate condition (C1R2) (Figure 2A-B).

When the second half of students received the race variation intervention, they also declined significantly more in the perception of genetic difference (χ^2 (1) = 21.7, p = 0.0001) and

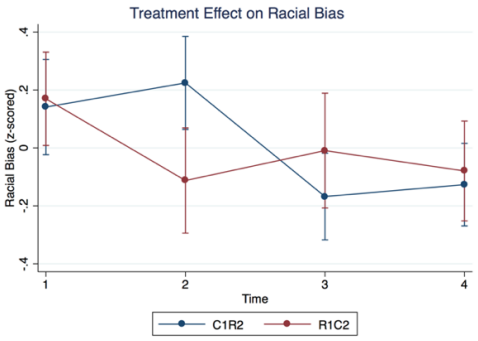
in racial bias (χ^2 (1) = 31.67, p = 0.0001) (ps < .05, after Bonferroni adjustment) compared to students who received the climate variation intervention second. Thus, our results are replicated in our cross-over design of RCT 1. Teaching these students about human genetic variation caused a reduction in racial bias (H_2) and the perception that races are genetically different (H_1).

Figure 2. Treatment effects for confirmatory analyses

A. Hypothesis One



B. Hypothesis Two

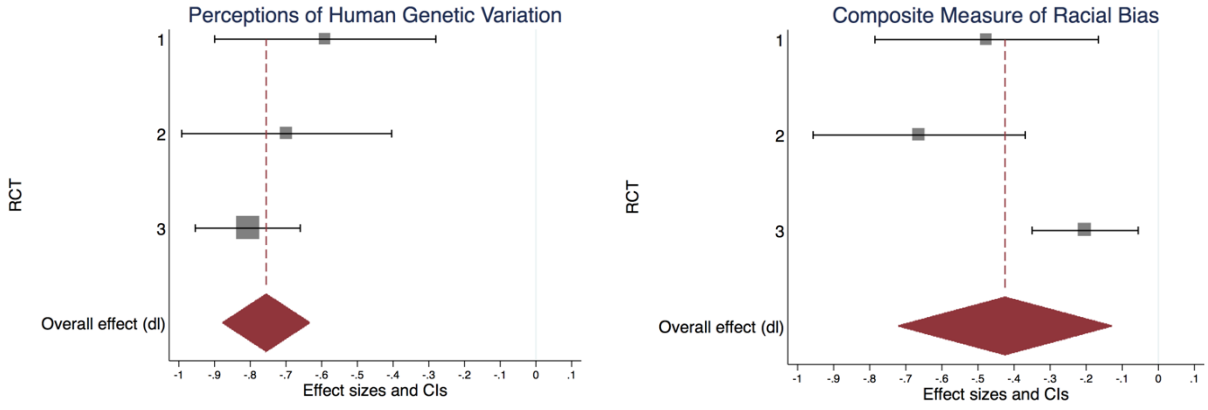


Notes: An important distinction to make in the interpretation of these graphs is when students in each experimental arm were treated with the race and human genetic variation intervention. Between timepoints 1 and 2, the R1C2 group received the race intervention and the C1R2 group received the climate intervention. Between timepoints 2 and 3 the R1C2 group received the climate intervention and the C1R2 group received the race intervention.

Reproducibility of H_1 and H_2 Findings in RCTs 2-3

Having established support for our hypotheses in RCT 1, we then explored the reproducibility of our findings by comparing the effect sizes of RCTs 1-3. Figure 3 represents the results of a DerSimonian and Laird (1986) random effects meta-analysis of the treatment effects of the race intervention on the variables used in each of the three RCTs.

Figure 3. Forest plot of random effects meta-analyses of RCTs 1-3



Note: The red diamond refers to the meta-analytic mean effect and confidence interval of the race intervention on each variable. Grey boxes and confidence intervals show the mean effect of the intervention in each RCT. Confidence intervals that do not overlap zero show a significant effect at p < 0.05. For RCT 1, we use the marginal treatment effects from time point 2 in Figure 2.

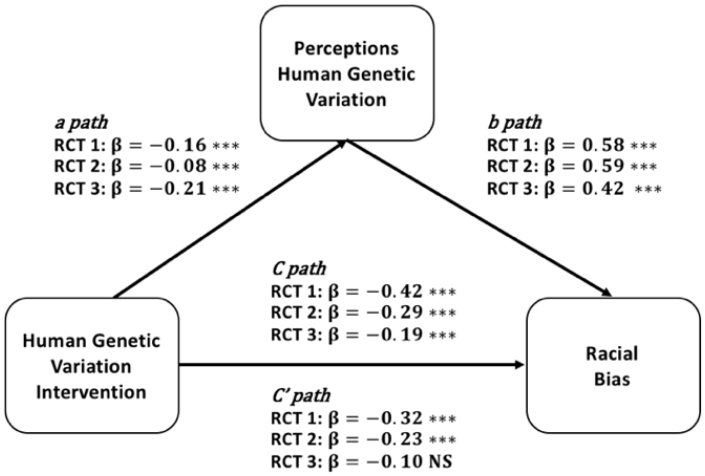
Perceptions of human genetic variation (H_1). As Figure 3 shows, RCTs 2 and 3 replicate the treatment effects on perceptions of human genetic variation observed in RCT 1 because each RCT produced the same effect magnitude (Cochrane's Q = 1.72, df = 2, p = 0.424, I^2 = 0%). Learning about human genetic variation results in a three-quarters standard deviation reduction in perceptions of human genetic variation (Cohen's d = -0.755, 95% CI [-0.876, -0.634]).

Racial bias (H_2). Figure 3 shows that across all three RCTs, learning about human genetic variation through our interventions caused a significant mean reduction in racial bias (Cohen's d = -0.425, 95% CI [-0.721, -0.129]). But, there was significant variability across studies in the magnitude of this effect (Cochrane's Q = 8.64, df = 2, p = 0.013, I^2 = 76.84%). The reductions in racial bias appear greater in RCT 2 (d = -0.663, 95% CI [-0.957, -0.369]) and RCT 1 (Cohen's d = -0.476, 95% CI [-0.786, -0.166]) compared to RCT 3 (d = -0.203, 95% CI [-0.350, -0.056]).

Support for Hypothesis 3 in RCTs 1-3.

Having found evidence in support of our first two hypotheses, we then explored whether the effect of the race intervention on racial bias was transmitted by changes in perceptions of human genetic variation (H_3). Figure 4 summarizes the results of these tests. In RCT 1, we find that 22.5% of the total reduction in racial bias caused by the race intervention is transmitted through perceptions of human genetic variation (bootstrapped indirect effect: β = -.093, 95% CI [-.201, -.028]). In RCT 2, we find that 18.6% of the total reduction in racial bias caused by the race intervention is transmitted through perceptions of human genetic variation (bootstrapped indirect effect: β = -.053, 95% CI [-.129, -.011]). In RCT 3, we find that 44.8% of the total reduction in racial bias caused by the race intervention is transmitted through perceptions of human genetic variation (bootstrapped indirect effect: β = -.086, 95% CI [-.148, -.032]). As these effects are not significantly different, learning about genetic variation appears to reduce racial bias by changing how adolescents perceive human genetic variation.

Figure 4. Exploratory tests of the mediated treatment effect (hypothesis 3)



Notes: *** p < 0.001, NS = not statistically significant. For RCT 1 we tested for mediation using the data only from the first half of the cross-over trial at time point 2.

To understand how this occurred, we re-ran each mediation model using a disaggregated form of the perceptions of human genetic variation instrument, using only the between group items or only the within group items. In RCTs 1-3 we found no evidence of an indirect effect using the within group items ($-.030 \leq \beta s \leq -.00049$, $ps \geq 0.28$) but we did find evidence of an indirect effect using the between group items ($-.126 \leq \beta s \leq -.052$, $ps \leq 0.019$). Therefore, learning about human genetic variation appears to reduce racial bias by affecting how students perceive the discreteness of racial categories rather than how students perceive the uniformity of same race individuals.

Duration of Effects in RCT 3

Having found evidence in support of the humane genetics education hypothesis we then explored whether the effects on racial bias persisted over time. We collected a delayed post measurement of both dependent variables three weeks after treatment at our California school site (RCT 3, n = 266). Using ordinary least squares regression, we found that students in the race condition still perceived significantly less variability between US census races than students assigned to the climate condition ($\beta = -.09$, SE = .032, $p = 0.008$) and also had significantly lower levels of racial bias ($\beta = -.26$, SE = .095, $p = 0.006$). We then conducted a time-lagged mediation model to explore if changes in perceptions of genetic variation measured immediately after treatment mediated reductions in racial bias three weeks later. We found that 32% of the treatment effect on racial bias in the delayed post was mediated by changes in how students perceived genetic variation immediately after they were treated three weeks earlier (Indirect effect $\beta = -.078$, SE = .039, $p = 0.048$). Since these effects were not reversed, there is no evidence that the race intervention backfired due to a 'Lewandowsky effect'. In the supplemental we support this claim further by using a multi-level Bayesian analysis of our data.

Discussion

Across three RCTs we demonstrate that when students were taught scientifically accurate information about human genetic variation, it reduced their perception that races are genetically different and their racial bias. These findings establish proof of concept for the hypothesis that human genetics education can be used to create a more humane society (i.e. less prejudiced) by influencing social cognition. We predicted these effects on the basis of prior research which suggested that learning about human genetic variation could reduce racial bias either by undermining the belief that same race people are genetically uniform and/or the idea that racial groups are discrete and non-overlapping categories. Our mediation analyses suggest that racial bias was reduced by our interventions primarily because individuals perceived less between group variation rather than more within group variation. However, this finding does not mean that learning about within group variation is unimportant as the two forms of variation are proportionally related. A learner must make sense of within group variation if they are to understand between group variation or the idea that racial groups are genetically alike in their variable DNA because they share similar sets of within group differences (Figure 1). Constructing an understanding of these concepts together reduced racial bias.

To elaborate further, when people are led to believe that scientists can distinguish one group from another based on traits that are similar within a group and different between groups, they are more likely to search for an underlying essence to differentiate ingroups from outgroups (Yzerbyt et al., 2001). Increased belief in an inherent essence then leads individuals to evaluate

ingroups more favorably than outgroups (Yzerbyt et al., 2001) and to categorize people into more discrete racial groups (Chao, Hong, & Chiu, 2013). In turn, a belief in racial discreteness has been found to lead people to stereotype outgroups more strongly because it causes them to perceive illusory correlations between outgroups and traits (Yzerbyt et al., 2001). Arguably, we ran this mechanism in reverse. By helping learners construct an understanding of genetic variation within and between groups, our interventions changed how students perceived between group variation, which reduced racial bias by undermining belief in racial discreteness.

An ambiguity in our findings is how instructional delivery and sampling characteristics interact to affect their replicability. We used two different kinds of interventions that differed because they involved or did not involve group-based discussions. We then tested these two different instructional approaches on two different populations (adolescent & adult). The invariance of the treatment effects on perceptions of human genetic variation across all three RCTs suggest that a humane genetics education can reduce the perception of genetic variation between US census races in samples of similar demographics drawn from culturally similar schools. These changes in perceptions of human genetic variation should be possible if teachers use curriculum and instruction that is aligned with the learning theory and ideas of a humane genetics education.

Whether teachers can reproduce the effect on racial bias is somewhat less certain than whether they can reproduce the effect on perceptions of human genetic variation. Our longer group-based intervention and shorter computer-based intervention both caused significant reductions in racial bias among adolescent learners (RCTs 1 & 3) but there was significant treatment effect variability in the magnitude of these reductions. For instance, the reductions in RCT 1 were greater than the reductions in RCT 3. Aboud and Fenwick (1999) argue that school-based debiasing interventions work better when high and low prejudice students talk together. They found that when high prejudice students made stereotypical claims about a group, low prejudice students would propose counter-examples to the stereotype during classroom discussions. In this zone of proximal development, high prejudice students revised their thinking through social interaction. Students in RCT 1 argued about racial difference using genetic evidence in small groups whereas students in RCTs 2 and 3 did not. Therefore, the cognitive work of learners during a humane genetics education arguably needs to be mediated by teachers and situated within group discussions for many days to produce a substantial reduction in racial bias. A computer-based intervention could be unsuccessful at reducing racial bias or even backfire in some schools.

Social norms surrounding the acceptability of racist speech at the schools we sampled could also be an enabling factor for the reductions in racial bias we observed. Monteiro et al. (2009) have found that older white children from working class families in Portugal impede blatantly racist thoughts when they sense anti-racist norms in their cultural environment. But they express these racially biased thoughts when they believe it is socially-acceptable. Since we were only able to sample school districts with research priorities aligned with reducing stereotyping, it is possible that we only observed reductions in racial bias because the students we sampled were changing their beliefs to act in accordance with the social norms of their school culture. However, experiments conducted in US public high schools rather than Portuguese psychology laboratories have found that social norms strongly affect discriminatory behavior but have no impact on beliefs (Paluck & Shepherd, 2012). Furthermore, there are good reasons to think that the racial bias effects we observed were not due to social desirability bias.

First, we demonstrated that reductions in racial bias were transmitted by changes in perceptions of human genetic variation induced by our interventions. Approximately 18%-44% of the total reduction in racial bias was significantly associated with changes in perceptions of human

genetic variation. Arguably, then, at least 18%-44% of the changes in racial bias we observed in each RCT was due to learning and not to social desirability bias. This implies that the remaining reductions in racial bias we observed actually are attributable to social desirability bias. Yet, in the supplemental we report additional findings from RCT 3 to demonstrate that the intervention effects on racial bias were not moderated by social desirability bias. Studies could estimate the impact of social desirability bias through a factorial design that randomizes a humane genetics education along with the presence or absence of clear anti-racist norms. If the effects of a humane genetics education are larger when anti-racist norms are present, and if these effects are not mediated by changes in how students perceive human genetic variation, then this would undercut our argument that learning about human genetic variation reduces racial bias.

A related alternative explanation for our findings is an experimenter demand effect, which is when people change their behavior in a study to conform to the subjective interpretations of the purpose of a study (Zizzo, 2008). Yet, experimenter demand effects are part and parcel of a humane genetics education. Schooling is a cultural endeavor where students become encultured into a group by learning how to be competent users of the group's conceptual tools (Brown, Collins, & Duguid, 1989; Lave & Wenger, 1991). Our interventions educated students about the conceptual tools that geneticists have used to combat racism in the culture of science. Thus, a student's subjective interpretation of the purpose of a humane genetics education is part of the power of it. We want learners to identify with the scientists who have challenged scientific racism and we want them to apply the tools of genetics to combat prejudice in their social worlds. The ultimate purpose of a humane genetics education is to identify with the social norms of anti-racist scientific culture.

Of course, our studies cannot evaluate whether our interventions achieved that ultimate purpose. Our findings say nothing about the domain transfer of these effects or their duration beyond three weeks in the schools we sampled. We have no idea how the knowledge learned in these interventions interacted with students' social goals, values, and other beliefs to affect their behavior in social domains outside of the biology classroom when students feel that race is salient for judging social behavior. Additionally, longitudinal studies may find that these effects wane or reverse after three weeks because students misremember what they learned (Lewandowsky et al., 2012). We found no evidence of such a 'Lewandowsky effect' at our California school site, but this does not mean that our intervention did not backfire at other sites or over longer time periods. In the supplemental we address the issue of intervention backfiring in more depth through a Bayesian analysis of our data. There we argue that the probability of intervention backfiring is relatively small compared to the probability of a clinically significant reduction in racial bias.

Implications

In closing, the interaction between cultural context and a humane genetics education needs to be explored before generalizations about the impact of this learning on social behavior are warranted. Nevertheless, our findings provide an initial proof of concept of the social cognitive consequences of a humane genetics education. By helping learners construct accurate perceptions of human genetic variation we reduced racial bias. Previous studies have found that repeated exposure to racial terminology in the biology curriculum can significantly increase belief in genetic essentialism because it leads students to perceive too much genetic variation between races (Donovan, 2017). Since our study is the first to show that the opposite effect is also possible, these results tentatively suggest that biology education is a lever that affects the development of racial bias, for better or worse, when it affects how students perceive genetic variation between racially-

defined populations. When biology education increases the perception that races are genetically different it can increase racial bias, and when it reduces such perceptions, it can decrease racial bias. If this mechanism is correct and generalizable, then the implications of our results for biology education in racially-diverse democracies could not be more apparent in an era when white nationalism is gaining political power in the US and in Europe (Jacobs, 2015).

Racialist political movements use biological essentialism to justify the oppression of minority groups (Omi & Winant, 1994). For example, discredited ideas about the biology of race were used by opponents of *Brown vs. Board of Education* in arguments to overturn it (Jackson Jr. & Depew, 2017). That opposition was halted, in part, because other scientists used population genetics to discredit the essentialist assumptions upon which opposition to school integration rested (Jackson Jr. & Depew, 2017). Ideas about human genetic variation have been and will continue to be important to policy debates about racial inequality (Byrd & Ray, 2015). Belief in genetic essentialism is still predictive of opposition to racially-ameliorative policies in white (Byrd & Ray, 2015) and non-white adults in the US (Soylu Yalcinkaya et al., 2017).

Since a major purpose of science education is to help learners understand the science behind public policy debates (National Research Council, 2012; J. Osborne & Monk, 2000) we cannot afford to ignore issues of race when we teach about human genetic variation. Unfortunately, current evidence suggests that when youth in predominantly white schools read textbooks describing the prevalence of genetic diseases in racial groups, it significantly increases their belief in genetic essentialism (Donovan, 2014, 2016, 2017). This increase reduces support for policies that redress racial inequality in education probably because it increases students' tendencies to explain racial disparities in education with genes (Donovan, 2016, 2017). If a 21st century biology education is to prevent the scientific racism of the past as we move into the genomic future, then it will need to offer youth a more humane genetics education than it currently does.

Such a proposition raises questions about the politics of official knowledge (Apple, 1993), particularly who is empowered to teach a humane genetics education? Regarding this question, it is important to point out that, depending on which country is sampled, Castéra and Clément (2014) estimate that 3-62% of biology teachers in European, South-American, African, or Middle-Eastern countries believe that, "Ethnic groups are genetically different and that is why some are superior to others" (e.g. 3% in France; 18% in Senegal; 34% in Poland; 62% in Lebanon). Even semi-nomadic herders in Mongolia exhibit biological essentialist beliefs about groups (Gil-White, 2001) Essentialist thinking appears at a non-zero prevalence in every human population we have gone looking for it (Henrich, Heine, & Norenzayan, 2010). In the US, representative studies estimate that 4% of preK-12 educators believe that racial inequalities are mainly due to a lower inborn potential to learn among African-Americans (Quinn, 2017).

Arguably, a humane genetics education could produce inhumane outcomes in the hands of such teachers, and there could be some of these teachers anywhere that genetics is taught. However, we demonstrated in RCT 2 that genetic essentialism can be reduced, at least temporarily, in a sample of US adults through our intervention. Thus, it is possible that educating teachers about human genetic variation to challenge faulty assumptions about racial difference could reduce racial biases in the teaching population. Moreover, there are plenty of teachers who do not believe that racial inequalities are due to genetics and these educators could be the first to be empowered to teach a more humane genetics education. Science educators have begun to outline the subject matter knowledge that teachers must possess to teach about genetic variation to reduce racism (Donovan, 2015a) and they have outlined curriculum frameworks for that teaching (J. Beckwith et al., 2017; Donovan, 2015b; Hubbard, 2017a). These frameworks can orient the interested

educator toward a more humane genetics education. But, more research needs to be done to chart a path through what is certain to be complex educational terrain. Those interested in navigating that path should know that when science shapes our perceptions of human genetic variation it can indirectly shape our prejudices as well. Therefore, teaching and learning about human genetic variation is not socially neutral.

References

Aboud, F. E., & Fenwick, V. (1999). Exploring and Evaluating School-Based Interventions to Reduce Prejudice. *Journal of Social Issues*, *55*(4), 767–786.

Alvermann, D. E., Smith, L. C., & Readence, J. E. (1985). Prior Knowledge Activation and the Comprehension of Compatible and Incompatible Text. *Reading Research Quarterly*, *20*(4), 420–436. https://doi.org/10.2307/747852

Andreychik, M. R., & Gill, M. J. (2014). Do natural kind beliefs about social groups contribute to prejudice? Distinguishing bio-somatic essentialism from bio-behavioral essentialism, and both of these from entitativity. *Group Processes & Intergroup Relations*, *18*(4), 454–474.

Apple, M. (1993). The politics of official knowledge: Does a national curriculum make sense? *The Teachers College Record*, *95*(2), 222–241.

Bastian, B., & Haslam, N. (2006). Psychological essentialism and stereotype endorsement. *Journal of Experimental Social Psychology*, *42*(2), 228–235. https://doi.org/10.1016/j.jesp.2005.03.003

Beckwith, J., Bergman, K., Carson, M., Doerr, T., Geller, L., Pierce, R., … Murray, A. V. (2017). Using Dialogues to Explore Genetics, Ancestry, and Race. *The American Biology Teacher*, *79*(7), 525–537.

Beckwith, J. R. (2009). *Making genes, making waves: A social activist in science*. Cambridge, Mass.: Harvard University Press. Retrieved from https://books.google.com/books?hl=en&lr=&id=KwlmBMcW9bcC&oi=fnd&pg=PP6&dq=making+genes+making+waves&ots=mcLAKf9l1D&sig=YSfk4HjS1QUflqjwBMYMINPM7n8

Beckwith, J. R., & Pierce, R. (2018). Genes and Human Behavior: Ethical Implications. In R. T. Gerlai (Ed.), *Molecular-Genetic and Statistical Techniques for Behavioral and Neural Research* (pp. 599–622). Academic Press: Elselvier.

Becquet, C., Patterson, N., Stone, A. C., Przeworski, M., & Reich, D. (2007). Genetic structure of chimpanzee populations. *PLoS Genet*, *3*(4), e66.

Bigler, R. S., & Liben, L. S. (2007). Developmental Intergroup Theory: Explaining and Reducing Children's Social Stereotyping and Prejudice. *Current Directions in Psychological Science*, *16*(3), 162–166. https://doi.org/10.1111/j.1467-8721.2007.00496.x

Boas, F. (1911). *The Mind of Primitive Man*. New York, NY: Macmillan.

Boas, F. (1928). *Anthropology and Modern Life*. New York, NY: W. W. Norton & Company.

Bowling, B. V., Acra, E. E., Wang, L., Myers, M. F., Dean, G. E., Markle, G. C., … Huether, C. A. (2008). Development and Evaluation of a Genetics Literacy Assessment Instrument for Undergraduates. *Genetics*, *178*(1), 15–22. https://doi.org/10.1534/genetics.107.079533

Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, *18*(1), 32–42.

Brueckner, H., Morning, A., & Nelson, A. (2005). The expression of biological concepts of race. In *Annual Meeting of the American Sociological Association, Philadelphia, PA*.

Byrd, W. C., & Ray, V. E. (2015). Ultimate Attribution in the Genetic Era: White Support for Genetic Explanations of Racial Difference and Policies. *The ANNALS of the American Academy of Political and Social Science*, *661*(1), 212–235. https://doi.org/10.1177/0002716215587887

Carter, P. L. (2012). *Stubborn Roots: Cultivating Cultural Flexibility and Equity in US and South African Schools*. Oxford University Press, Incorporated.

Caspi, A., & Moffitt, T. E. (2006). Gene–environment interactions in psychiatry: joining forces with neuroscience. *Nature Reviews Neuroscience*, *7*(7), 583–590.

Castéra, J., & Clément, P. (2014). Teachers' conceptions about the genetic determinism of human behaviour: A survey in 23 countries. *Science & Education*, *23*(2), 417–443.

Chao, M. M., Hong, Y., & Chiu, C. (2013). Essentializing race: Its implications on racial categorization. *Journal of Personality and Social Psychology*, *104*(4), 619–634.

Chinn, C. A., & Malhotra, B. A. (2002). Children's responses to anomalous scientific data: How is conceptual change impeded? *Journal of Educational Psychology*, *94*(2), 327–343. https://doi.org/10.1037/0022-0663.94.2.327

Cimpian, A., & Salomon, E. (2014). The inherence heuristic: An intuitive means of making sense of the world, and a potential precursor to psychological essentialism. *Behavioral and Brain Sciences*, *37*(05), 461–480. https://doi.org/10.1017/S0140525X13002197

Coop, G., Eisen, M. B., Nielsen, R., Przeworski, M., & Rosenberg, N. A. (2014). Letters: "A Troublesome Inheritance" | Stanford Center for Computational, Evolutionary, and Human Genomics. Retrieved June 15, 2018, from https://cehg.stanford.edu/letter-from-population-geneticists

Crawford, N. G., Kelly, D. E., Hansen, M. E. B., Beltrame, M. H., Fan, S., Bowman, S. L., … Tishkoff, S. A. (2017). Loci associated with skin pigmentation identified in African populations. *Science*, *358*(6365), eaan8433. https://doi.org/10.1126/science.aan8433

Dar-Nimrod, I., & Heine, S. J. (2011). Genetic essentialism: On the deceptive determinism of DNA. *Psychological Bulletin*, *137*(5), 800–818. https://doi.org/10.1037/a0021860

DerSimonian, R., & Laird, N. (1986). Meta-analysis in clinical trials. *Controlled Clinical Trials*, *7*(3), 177–188.

Devlin, B., Daniels, M., & Roeder, K. (1997). The heritability of IQ. *Nature*, *388*, 908–912.

Dobzhansky, T. (1937). *Genetics and the Origin of Species* (Vol. 11). Columbia university press.

Donovan, B. M. (2014). Playing with fire? The impact of the hidden curriculum in school genetics on essentialist conceptions of race. *Journal of Research in Science Teaching*, *51*(4), 462–496. https://doi.org/10.1002/tea.21138

Donovan, B. M. (2015a). Putting humanity back into the teaching of human biology. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *52*, 65–75. https://doi.org/10.1016/j.shpsc.2015.01.011

Donovan, B. M. (2015b). Reclaiming Race as a Topic of the U.S. Biology Textbook Curriculum. *Science Education*, *99*(6), 1092–1117. https://doi.org/10.1002/sce.21173

Donovan, B. M. (2016). Framing the Genetics Curriculum for Social Justice: An Experimental Exploration of How the Biology Curriculum Influences Beliefs About Racial Difference. *Science Education*, *100*(3), 586–616. https://doi.org/10.1002/sce.21221

Donovan, B. M. (2017). Learned inequality: Racial labels in the biology curriculum can affect the development of racial prejudice. *Journal of Research in Science Teaching*, *54*(3), 379–411. https://doi.org/10.1002/tea.21370

Edge, M. D., & Rosenberg, N. A. (2015). Implications of the apportionment of human genetic diversity for the apportionment of human phenotypic diversity. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*. https://doi.org/10.1016/j.shpsc.2014.12.005

Emmons, N. A., & Kelemen, D. A. (2015). Young children's acceptance of within-species variation: Implications for essentialism and teaching evolution. *Journal of Experimental Child Psychology*, *139*, 148–160. https://doi.org/10.1016/j.jecp.2015.05.011

Engberg, M. E. (2004). Improving Intergroup Relations in Higher Education: A Critical Examination of the Influence of Educational Interventions on Racial Bias. *Review of Educational Research*, *74*(4), 473–524. https://doi.org/10.3102/00346543074004473

Enoch, M.-A. (2012). The Influence of Gene–Environment Interactions on the Development of Alcoholism and Drug Dependence. *Current Psychiatry Reports*, *14*(2), 150–158. https://doi.org/10.1007/s11920-011-0252-9

Evans, E. M., Spiegel, A. N., Gram, W., Frazier, B. N., Tare, M., Thompson, S., & Diamond, J. (2010). A conceptual guide to natural history museum visitors' understanding of evolution. *Journal of Research in Science Teaching*, *47*(3), 326–353. https://doi.org/10.1002/tea.20337

Feldman, M., & Lewontin, R. (1975). The heritability hang-up. *Science*, *190*(4220), 1163–1168. https://doi.org/10.1126/science.1198102

Feldman, Marcus. (2014). Echoes of the Past: Hereditarianism and A Troublesome Inheritance. *PLoS Genetics*, *10*(12), e1004817. https://doi.org/10.1371/journal.pgen.1004817

Flynn, J. R. (1999). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, *54*(1), 5–20.

Ford, M. J. (2012). A Dialogic Account of Sense-Making in Scientific Argumentation and Reasoning. *Cognition and Instruction*, *30*(3), 207–245. https://doi.org/10.1080/07370008.2012.689383

Foster, M. W. (2009). Looking for race in all the wrong places: analyzing the lack of productivity in the ongoing debate about race and genetics. *Human Genetics*, *126*(3), 355–362. https://doi.org/10.1007/s00439-009-0674-1

Gelman, S. A. (2004). Psychological essentialism in children. *Trends in Cognitive Sciences*, *8*(9), 404–409.

Gil-White, F. J. (2001). Are Ethnic Groups Biological "Species" to the Human Brain? Essentialism in Our Cognition of Some Social Categories. *Current Anthropology*, *42*(4), 515–553. https://doi.org/10.1086/321802

Gould, S. J. (1996). *The mismeasure of man*. WW Norton & Company.

Graves, J. L. (2015). Great Is Their Sin: Biological Determinism in the Age of Genomics. *The ANNALS of the American Academy of Political and Social Science*, *661*(1), 24–50. https://doi.org/10.1177/0002716215586558

Guzzetti, B. J., Snyder, T. E., Glass, G. V., & Gamas, W. S. (1993). Promoting Conceptual Change in Science: A Comparative Meta-Analysis of Instructional Interventions from

Reading Education and Science Education. *Reading Research Quarterly*, *28*(2), 116–159. https://doi.org/10.2307/747886

Haberstick, B. C., Lessem, J. M., Hewitt, J. K., Smolen, A., Hopfer, C. J., Halpern, C. T., … Mullan Harris, K. (2014). MAOA Genotype, Childhood Maltreatment, and Their Interaction in the Etiology of Adult Antisocial Behaviors. *Biological Psychiatry*, *75*(1), 25–30. https://doi.org/10.1016/j.biopsych.2013.03.028

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*(2–3), 61–83. https://doi.org/10.1017/S0140525X0999152X

Hubbard, A. R. (2017a). Teaching Race (Bioculturally) Matters: A Visual Approach for College Biology Courses. *The American Biology Teacher*, *79*(7), 516–524. https://doi.org/10.1525/abt.2017.79.7.516

Hubbard, A. R. (2017b). Testing Common Misconceptions about the Nature of Human Racial Variation. *The American Biology Teacher*, *79*(7), 538–543. https://doi.org/10.1525/abt.2017.79.7.538

Hunley, K. L., Cabana, G. S., & Long, J. C. (2016). The apportionment of human diversity revisited. *American Journal of Physical Anthropology*, *160*(4), 561–569. https://doi.org/10.1002/ajpa.22899

Jablonski, N. G., & Chaplin, G. (2010). Human skin pigmentation as an adaptation to UV radiation. *Proceedings of the National Academy of Sciences*, *107*(Supplement 2), 8962–8968.

Jackson Jr., J. P., & Depew, D. J. (2017). *Darwinism, Democracy, and Race: American Anthropology and Evolutionary Biology in the Twentieth Century*. New York, NY: Routledge.

Jacobs, T. S. (2015). *White Pride World Wide: A Contemporary History of Online, International Right-wing Extremist Ideology and Culture*. University of Oslo. Retrieved from https://www.duo.uio.no/handle/10852/45238

Jayaratne, T. E., Ybarra, O., Sheldon, J. P., Brown, T. N., Feldbaum, M., Pfeffer, C. A., & Petty, E. M. (2006). White Americans' Genetic Lay Theories of Race Differences and Sexual Orientation: Their Relationship with Prejudice toward Blacks, and Gay Men and Lesbians. *Group Processes & Intergroup Relations*, *9*(1), 77–94. https://doi.org/10.1177/1368430206059863

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

Kahan, B. C., & Morris, T. P. (2013). Assessing potential sources of clustering in individually randomised trials. *BMC Medical Research Methodology*, *13*(58), 1–9.

Kaplan, J. M., & Winther, R. G. (2013). Prisoners of Abstraction? The Theory and Measure of Genetic Variation, and the Very Concept of "Race." *Biological Theory*, *7*(4), 401–412.

Kaplan, Jonathan Michael, & Winther, R. G. (2013). Realism, Antirealism, and Conventionalism about'Race'. Retrieved from http://philpapers.org/rec/KAPRAA

Keller, M. C. (2014). Gene × Environment Interaction Studies Have Not Properly Controlled for Potential Confounders: The Problem and the (Simple) Solution. *Biological Psychiatry*, *75*(1), 18–24. https://doi.org/10.1016/j.biopsych.2013.09.006

Kimel, S. Y., Huesmann, R., Kunst, J. R., & Halperin, E. (2016). Living in a Genetic World How Learning About Interethnic Genetic Similarities and Differences Affects Peace and Conflict. *Personality and Social Psychology Bulletin*, *42*(5), 688–700.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge university press. Retrieved from http://books.google.com/books?hl=en&lr=&id=CAVIOrW3vYAC&oi=fnd&pg=PA11&dq=lave+1991&ots=OAqzwq0FFk&sig=aS0_7z3isT6rYY3MdFDsaFo5nxw

Levy, S. R., Stroessner, S. J., & Dweck, C. S. (1998). Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology*, *74*(6), 1421–1436.

Lewandowsky, S., Ecker, U. K. H., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and Its Correction: Continued Influence and Successful Debiasing. *Psychological Science in the Public Interest*, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lewontin, R. (1996). *Biology as ideology: The doctrine of DNA*. House of Anansi.

Lewontin, R. C. (1972). The apportionment of human diversity. *Evolutionary Biology*, *6*, 381–398.

Livingstone, F. B., & Dobzhansky, T. (1962). On the Non-Existence of Human Races. *Current Anthropology*, *3*(3), 279–281.

Lynch, J., Bevan, J., Achter, P., Harris, T., & Condit, C. M. (2008). A preliminary study of how multiple exposures to messages about genetics impact on lay attitudes towards racial and genetic discrimination. *New Genetics and Society*, *27*(1), 43–56. https://doi.org/10.1080/14636770701843634

MacMahon, B. (1968). Gene-environment interaction in human disease. *Journal of Psychiatric Research*, *6*, 393–402.

Markus, H., & Moya, P. M. L. (2011). *Doing race: 21 essays for the 21st century*. WW Norton & Company. Retrieved from https://repository.library.georgetown.edu/handle/10822/548601

Monteiro, M. B., de França, D. X., & Rodrigues, R. (2009). The development of intergroup bias in childhood: How social norms can shape children's racial behaviours. *International Journal of Psychology*, *44*(1), 29–39. https://doi.org/10.1080/00207590802057910

Morning, A. J. (2011). *The Nature of Race: How Scientists Think and Teach about Human Difference*. London, England: University of California Press.

National Research Council. (2012). *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas*. National Academies Press.

NGSS. (2016). Standards by Topic | Next Generation Science Standards. Retrieved July 29, 2016, from http://www.nextgenscience.org/overview-topics

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

Omi, M., & Winant, H. (1994). *Racial Formation in the United States: From the 1960s to the 1990s*. Psychology Press.

Opfer, J. E., Nehm, R. H., & Ha, M. (2012). Cognitive foundations for science assessment design: Knowing what students know about evolution. *Journal of Research in Science Teaching*, *49*(6), 744–777. https://doi.org/10.1002/tea.21028

Osborne, J., & Monk, M. (2000). *Good practice in science teaching: What research has to say*. Berkshire, England: Open University Press.

Osborne, Jonathan, Donovan, B. M., Henderson, J. B., MacPherson, A. C., & Wild, A. (2016). *Arguing From Evidence in Middle School Science: 24 Activities for Productive Talk and*

Deeper Learning. Thousand Oaks: CA: Corwin Press. Retrieved from https://us.corwin.com/en-us/nam/arguing-from-evidence-in-middle-school-science/book250184#description

Paluck, E. L., & Shepherd, H. (2012). The salience of social referents: A field experiment on collective norms and harassment behavior in a school social network. *Journal of Personality and Social Psychology*, *103*(6), 899–915. https://doi.org/10.1037/a0030015

Parrott, R. L., Silk, K. J., Dillow, M. R., Krieger, J. L., Harris, T. M., & Condit, C. M. (2005). Development and validation of tools to assess genetic discrimination and genetically based racism. *Journal of the National Medical Association*, *97*(7), 980–990.

Pauker, K., Ambady, N., & Apfelbaum, E. P. (2010). Race Salience and Essentialist Thinking in Racial Stereotype Development. *Child Development*, *81*(6), 1799–1813.

Pauker, Kristin, Xu, Y., Williams, A., & Biddle, A. M. (n.d.). Race essentialism and social contextual differences in children's racial stereotyping. *Child Development. Advance Online Publication*. Retrieved from http://manoa.hawaii.edu/isplab/wp-content/uploads/2015/10/Pauker-Xu-Williams-Biddle-in-press_ChildDevelopment.pdf

Quinn, D. M. (2017). Racial Attitudes of PreK–12 and Postsecondary Educators: Descriptive Evidence From Nationally Representative Data. *Educational Researcher*, 0013189X17727270. https://doi.org/10.3102/0013189X17727270

Reich, D. (2018). *Who We Are and How We Got Here: Ancient DNA and the new science of the human past*. Oxford University Press.

Relethford, J. H. (2002). Apportionment of global human genetic diversity based on craniometrics and skin color. *American Journal of Physical Anthropology*, *118*(4), 393–398.

Relethford, John H. (1997). Hemispheric difference in human skin color. *American Journal of Physical Anthropology*, *104*(4), 449–457.

Rosenberg, N. A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., & Feldman, M. W. (2002). Genetic structure of human populations. *Science*, *298*(5602), 2381–2385.

Rosenberg, Noah A. (2011). A Population-Genetic Perspective on the Similarities and Differences among Worldwide Human Populations. *Human Biology*, *83*(6), 659–684. https://doi.org/10.3378/027.083.0601

Rosenberg, Noah A, & Edge, M. D. (in press). Genetic clusters and the race debates: a perspective from population genetics. In Q. Spencer (Ed.), *The Race Debates from Metaphysics to Medicine*. Oxford University Press, USA.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction*, *16*(4), 475–5223.

Shemwell, J., Chase, C. C., & Schwartz, D. L. (2015). Seeking the General Explanation: ATest of Inductive Activities for Learning and Transfer, *52*(1), 58–83.

Shtulman, A., & Schulz, L. (2008). The Relation Between Essentialist Beliefs and Evolutionary Reasoning. *Cognitive Science: A Multidisciplinary Journal*, *32*(6), 1049–1062. https://doi.org/10.1080/03640210801897864

Solomon, R. L., & Lessac, M. S. (1968). A control group design for experimental studies of developmental processes. *Psychological Bulletin*, *70*(3), 145–150.

Soylu Yalcinkaya, N., Estrada-Villalta, S., & Adams, G. (2017). The (Biological or Cultural) Essence of Essentialism: Implications for Policy Support among Dominant and

Subordinated Groups. *Frontiers in Psychology*, *8*. https://doi.org/10.3389/fpsyg.2017.00900

Tawa, J. (2016). Belief in Race as Biological: Early Life Influences, Intergroup Outcomes, and the Process of "Unlearning." *Race and Social Problems*, *8*(3), 244–255. https://doi.org/10.1007/s12552-016-9176-7

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, *14*(6), 623–628.

Van den Broek, P. (2010). Using texts in science education: Cognitive processes and knowledge representation. *Science*, *328*(5977), 453–456.

Wade, N. (2014). *A Troublesome Inheritance: Genes, Race and Human History*. Melbourne, Australia: Penguin Press. Retrieved from http://books.google.com/books?hl=en&lr=&id=hPp4AgAAQBAJ&oi=fnd&pg=PT4&dq=A+Troublesome+Inheritance:+Genes,+Race+and+Human+History&ots=hd-u9dAawd&sig=ae7nCqmc46Phu-5FFFHO0TonAys

Weiss, K. M. (2008). The good, the bad, and the ugli. *Evolutionary Anthropology: Issues, News, and Reviews*, *17*(3), 129–134. https://doi.org/10.1002/evan.20163

Weiss, M. J., Lockwood, J. R., & McCaffrey, D. F. (2016). Estimating the Standard Error of the Impact Estimator in Individually Randomized Trials With Clustering. *Journal of Research on Educational Effectiveness*, *9*(3), 421–444. https://doi.org/10.1080/19345747.2015.1086911

Williams, M. J., & Eberhardt, J. L. (2008). Biological conceptions of race and the motivation to cross racial boundaries. *Journal of Personality and Social Psychology*, *94*(6), 1033–1047.

Willinsky, J. (1998). *Learning to divide the world: Education at empire's end*. Minneapolis, MN: University of Minnesota Press.

Yudell, M., Roberts, D., DeSalle, R., & Tishkoff, S. (2016). Taking race out of human genetics. *Science*, *351*(6273), 564–565.

Yzerbyt, V., Corneille, O., & Estrada, C. (2001). The interplay of subjective essentialism and entitativity in the formation of stereotypes. *Personality and Social Psychology Review*, *5*(2), 141–155.

Zizzo, D. J. (2008). *Experimenter Demand Effects in Economic Experiments* (Social Science Research Network Discussion Paper) (p. 37).