Sociological Theory
http://stx.sagepub.com/

Does Genomics Challenge the Social Construction of Race?

Ann Morning

Sociological Theory 2014 32: 189
DOI: 10.1177/0735275114550881

The online version of this article can be found at:
http://stx.sagepub.com/content/32/3/189

Published by:
SAGE
http://www.sagepublications.com

On behalf of:
American Sociological Association

Additional services and information for *Sociological Theory* can be found at:

Email Alerts: http://stx.sagepub.com/cgi/alerts

Subscriptions: http://stx.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

>> Version of Record - Oct 10, 2014

What is This?

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

Article

AMERICAN SOCIOLOGICAL ASSOCIATION

Sociological Theory
2014, Vol. 32(3) 189–207
© American Sociological Association 2014
DOI: 10.1177/0735275114550881
stx.sagepub.com
SAGE

Does Genomics Challenge the Social Construction of Race?

Ann Morning[1]

Abstract

Shiao, Bode, Beyer, and Selvig argue that the theory of race as a social construct should be revisited in light of recent genetic research, which they interpret as demonstrating that human biological variation is patterned in "clinal classes" that are homologous to races. In this reply, I examine both their claims and the genetics literature they cite, concluding that not only does constructivist theory already accommodate the contemporary study of human biology, but few geneticists portray their work as bearing on race. Equally important, methods for statistically identifying DNA-based clusters within the human species are shaped by several design features that offer opportunities for the incorporation of cultural assumptions about difference. As a result, Shiao et al.'s theoretical distinction between social race and biological "clinal class" is empirically jeopardized by the fact that even our best attempts at objectively recording "natural" human groupings are socially conditioned.

Keywords

race, science, genetics, ethnicity, classification

Early on in "The Genomic Challenge to the Social Construction of Race," Shiao, Bode, Beyer, and Selvig (2012:67, 69) describe the social constructionist account of race as "lacking biological reality." As they see it, the constructivist perspective is "unnecessarily burdened with . . . a conception of human biological variation that is out of step with recent advances in genetic research" (Shiao et al. 2012:68). The authors' response is to propose a "solution" that reformulates racial constructionism by reconceptualizing the nature of racial (and ethnic) categorization (Shiao et al. 2012:69). Their "theoretical synthesis"

> accepts the existence of genetic clusters consistent with certain racial classifications as well as the validity of the genomic research that has identified the clusters, without diminishing the social character of their context, meaning, production, or consequences. We argue that the recent research in genetics demonstrates that certain racial, and also ethnic, categories have a biological basis in statistically discernible clusters of alleles rather than in the traditional notions of human races as arising from categorically

[1]New York University, New York, NY, USA

Corresponding Author:
Ann Morning, Department of Sociology, New York University, 295 Lafayette St., Rm. 4118, New York, NY 10012, USA.
Email: ann.morning@nyu.edu

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

> distinct ancestries or as possessing categorically unique essences. (Shiao et al. 2012:68-69)

In sum, Shiao et al. argue that current constructivist thinking about race (1) is not up to date, (2) should therefore be amended, and (3) can profitably be revised by drawing on claims from recent literature in human genetics that "certain racial, and also ethnic, categories have a biological basis" (Shiao et al. 2012:68).

I begin this reply to Shiao et al. (2012) by countering that constructionist theory is already quite capable of accounting for new (and not-so-new) claims about race and biology. Indeed, genetics-based claims about the "biological basis" of race offer a textbook case of social construction.

Underscoring this point, the focus of this article will be on examining Shiao et al.'s (2012) main argument that current research in human genetics warrants a reformulation of constructivist race theory. As it turns out, review of the scientific literature they cite reveals how deeply social the production of this strand of biological knowledge is. There are myriad ways in which "facts" about human genetic variation are shaped by analysts' assumptions and decisions: Choices about whose DNA to sample and which types of genetic data to analyze, as well as assumptions about how different populations are related to each other and decisions about what statistical techniques to employ, all bear on scientists' conclusions about the genetic "clusters" that ostensibly characterize our species. Moreover, they leave many openings for our widely shared beliefs about racial difference to filter in. Consequently, the examination of research reports in human genetics leads me to contest Shiao et al.'s claim that recent research has clearly demonstrated "that certain racial, and also ethnic, categories have a biological basis in statistically discernible clusters of alleles" (Shiao et al. 2012:68). Instead, statistically inferred human genetic clusters have a social basis that is entirely consistent with current constructivist thinking about race. Our depictions of the genetic structure of human populations are themselves so culturally conditioned that it would be a mistake to conclude that they represent objective biological measurements that are any less a human artifact than folk taxonomies.

Debates about how best to understand race—including how to define it and conceptualize its relationship to society and to biology—have far-reaching implications. Beliefs about the nature of race influence racial attitudes, practices, and policies (Morning 2011). In the past, essentialist notions of inherited biological racial characteristics underwrote eugenic thinking in the United States, leading to changes in immigration law that restricted the entry of people considered to be of inferior stock (Kevles 1985). Today, biological concepts of race affect practices ranging from medical diagnosis and treatment to forensic criminology and genealogy (Baer et al. 2013; Lee 2009; P. Martin et al. 2007; Nelson 2008; Ossorio and Duster 2005; Satel 2002). While some argue that such a model of race offers important medical attention and assistance to traditionally underserved communities of color (Burchard et al. 2003; Sarich and Miele 2004), others contend it obscures the social roots of inequality and thus detracts attention and support away from public policies—to bolster health care access or education, for example—that would address stratification (Duster 2006; Roberts 2011). In short, what genomic research can or cannot tell us about race is not simply a technical question of interest to a few scholars. Instead, it is at the heart of a much broader controversy about the place of race, particularly in American society, and the policies we should or should not adopt in the face of enduring inequalities.

THE ROLE OF BIOLOGY IN RACIAL CONSTRUCTIVISM

Shiao et al.'s (2012) perception of a disconnect between constructivist theory and contemporary genetic research hinges not only on their reading of the latter but, just as importantly,

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

on their interpretation of the former. A standard constructivist statement comes from the American Sociological Association (2003:7), which described race as "a social invention that changes as political, economic, and historical contexts change." Shiao et al. are sympathetic to this central view and its elaboration in Omi and Winant's (1986, 1994) theory of racial formation. From this basic emphasis on collective processes of construction, however, they deduce a corollary, namely, a "social constructionist refutation of race as possessing any biological reality" (Shiao et al. 2012:68).

In contrast, I submit that constructivist theory is not blind to the input that human biology can have in racial classification but that it rejects as empirically unwarranted the genetic determination of racial categories that Shiao et al.'s (2012) assertion of a "biological basis" (p. 68) for race at times implies. In this section, I will elaborate on the role that constructionism assigns to biology in the formulation of race categories—and the linkages it refutes—before turning in subsequent sections to a closer examination of Shiao et al.'s theoretical claims and their empirical merit.

To begin, constructivist theory accepts the idea that our assignment of individuals to racial groups is often based on our perceptions of their physical characteristics. As one introductory sociology textbook put it, a race denotes "a category of people who perceive themselves and are perceived by others as distinctive on the basis of certain biologically inherited traits" (Calhoun, Light, and Keller 2001).[1] To the extent that racial assignments incorporate information (alongside beliefs) about individuals' phenotypic characteristics or geographic origins, they can be said to be informed by (or correlated with) biology. If that is the "biological reality" of race that Shiao et al. (2012) wish to emphasize, it does not represent a novel contribution—much less a challenge—to the constructionist account.

Constructivist theories of race thus portray biological differences between racial groups as a byproduct of social classification practices, not as their driver or even a major contributor. In the constructionist perspective, categories are fashioned and given meaning according to sociopolitical imperatives—for example, who is native to a territory and who is a settler; who can be enslaved and who cannot; who is a member of the faithful, the ruling clan, or superior stock—and then we use phenotype and ancestry (among other factors; see Harris 1964) to guide our assignment of individuals to one racial group or another. In so doing—not to mention through subsequent generations' exposure to racially differentiated social and/or physical environments—socially constructed groupings can come to be roughly distinguished by biological differentials, for example, in skin color or frequency of a particular genetic variant or allele (Krieger 2005). It would be a mistake however to deduce from such post hoc distinctions that preexisting genetic or other biological differences had given rise to those groupings.

Racial constructionism offers several explanations for why a deterministic mapping of biology to race is indefensible. First, neither the historical origins nor the malleability of racial categories supports the claim that they "have a biological basis in statistically discernible clusters of alleles" (Shiao et al. 2012:68). The "white" category, for example, has expanded and contracted so dramatically over time—to include or exclude Jews, Irish people, Laplanders, Hispanics, South Asians, Middle Easterners, and Ethiopians, for example (Haney López 1996; Jacobson 1998; Marks 1995; Samhan 1999; Sanders 1969)—that it is hard to discern the force of "biological reality" at work. Similarly, the American "one-drop rule," which classifies as black all individuals of known or suspected African ancestry (including, for example, early immigrants from India; see Helweg and Helweg 1990), obviously runs against the grain of the inheritance of genetic traits, which an individual acquires from both mother and father, not simply from a black parent to the exclusion of a white one.

Second, constructivist theory reminds us that the phenotypic information we use to make racial assignments does not represent a comprehensive or representative summary of

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

individuals' biological data; instead, we give great weight to a few traits, such as skin color, hair texture, and eye shape. In contrast, other characteristics rooted in DNA—for example, blood type, resistance to malaria, or lactose intolerance—do not factor into our racial classifications. If they did, they would produce a different taxonomy of races, since they are not concordant with the somatic traits mentioned above (American Anthropological Association 1998; Marks 1995). Winant (2001:317) alerts us to the fundamentally social nature of the selection of physical traits we privilege in our racial conceptualization and classification when he defines race as

> a concept that signifies and symbolizes sociopolitical conflicts and interests in reference to different types of human bodies. Although the concept of race appeals to biologically based human characteristics (so-called phenotypes) selection of these particular human features for purposes of racial signification is always and necessarily a social and historical process.

In other words, the selection of a few physical traits that are not concordant with many others as criteria for racial classification does not inspire confidence in the notion that naturally occurring, statistically meaningful, and objectively or comprehensively constituted clusters within the human species are what gave rise to today's racial categories.

MODELING THE RELATIONSHIP BETWEEN GENES AND RACE

Having considered Shiao et al.'s (2012) claims about constructionist theory, I turn now to the "solution" they propose for better conceptualizing the nature of race. The authors seem to wish to reformulate the relationship between biology and race in a way that neither simply reiterates the widespread observation that people often interpret others' physical characteristics when classifying by race nor is so extremely deterministic as to maintain that our current categories—for example, those used by the U.S. Census (Humes, Jones, and Ramirez 2011)—are dictated by naturally distinct genetic clusters in the human species. What is the middle ground they seek, then, and is it empirically validated?

Central to Shiao et al.'s (2012) model of the relationship between biology and race is a theoretical distinction between biologically rooted versus socially constructed groupings. In particular, they advocate "replacing the refutation of biology in racial constructionism with a version of the feminist distinction between biological sex and socially constructed gender (Rubin 1975)" (Shiao et al. 2012:72). Accordingly, the authors propose "to conceptualize racial and ethnic categorization as the social perception of biological ancestry" (Shiao et al. 2012:77). In other words, they distinguish objective, natural, biological "measurement of ancestry" (Shiao et al. 2012:79) from subjective, constructed, social "perception of ancestry."

Equally important to Shiao et al.'s (2012) theoretical agenda is the introduction of the concept of "clinal class" (p. 72) as the measure of biological ancestry that is the counterpart of social races. Depicting these classes as "the lumps in otherwise continuous genetic variation" (Shiao et al. 2012:69), the authors seek to reconcile two features of human biology. On one hand, genetic and phenotypic traits generally follow a "clinal" or gradated pattern of variation across space; rather than displaying sharp differences between neighboring populations, small, gradual changes are the rule. On the other hand, many researchers hypothesize that human biological variation is not always so smooth but rather may also present recognizable "lumps" or "clusters" of individuals who are distinct from their neighbors. Shiao et al. marry cline and cluster as follows: "As already noted, the concept of subspecies, human

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

or otherwise, does not require categorical differences between populations. Instead, clinal classes assume a common evolutionary history, possess extensive genetic similarities, and coexist with clinal variations both within each class and across classes" (Shiao et al. 2012:72).

As a newly coined term, "clinal class" is at present a theoretical concept that is not used in genetics literature, where instead "clines" and "clusters" are the usual references in work on human population structure. Without a body of empirical research that illustrates or elaborates on the term, it is difficult to assess whether and/or how clinal classes relate to "social" racial and ethnic categories. Yet it is precisely this hypothesized relationship that is at the heart of Shiao et al.'s (2012) argument regarding the biological basis of racial classification.

The connection between clinal class and race is not immediately obvious; Shiao et al. (2012) often describe them as "homologous" to another, a formulation that could admit many interpretations. In fact, the authors envision multiple potential configurations in which clinal classes overlap or not with race categories. In one scenario, "select clinal classes, and not others, become defined as salient racial/ethnic categories" (Shiao et al. 2012:73); in another, "racial/ethnic classification can be accurate in terms of making categories that are homologous to clinal classes. . . . And vice versa, racial /ethnic classifications may be accurately associated with empirical group averages but not be homologous to verifiable clinal classes when the perceived ancestries have no biological basis" (Shiao et al. 2012:79). In other words, clinal classes do not always translate into races, and races do not always correspond to clinal classes. Instead, the association between them can be strong or weak:

> Clinal classes may be the natural boundaries for racial/ethnic classifications that vary simply according to the patterns of who has had contact with whom. Similarly, clinal classes may serve as the objective variation on which racial/ethnic classifications make variable divisions, similar to the color perception of the light spectrum, that is, whether blue and green are considered shades of the same color or distinct colors and by what terms "blue" and "green" are called. Alternatively, clinal classes may provide an objective but less pressing variation from which racial/ethnic classifications make selections and specifications, for example, giving some clinal classes more cultural "weight" than others (Danna 2011). Most distinctly, clinal classes may provide no constraint on racial/ethnic classifications where socially recognized ancestries have no basis in verifiable clinal classes. (Shiao et al. 2012:84)

In short, Shiao et al. posit a spectrum along which genetic clinal classes may contribute more or less to the social construction of race and ethnicity. Without any empirical research on clinal classes from which to draw, however, the authors cannot identify or substantiate concretely which racial or ethnic categories are strongly homologous to such classes and which are not.

Why and under what conditions clinal classes might or might not be mirrored in racial categories are not explained in "The Genomic Challenge to the Social Construction of Race." Constructivist theory would suggest their alignment would be "hit and miss"—hardly the rule—because racial categories are constructed on the basis of cultural calculations of social and political similarity that are not dependent on biological traits. And when the two are more closely related at present, this may be due not to distinctive biological group characteristics that informed the original classification scheme but rather to generations of stratifying social practices that have tended to genetically isolate the people racially labeled in a certain way.

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

Finally, another important potential explanation for the occasional alignment of hypothesized clinal classes with socially constructed races remains, namely, that our measures of "biological ancestry" are molded by the same cultural beliefs about human difference that fuel our social classification schemes. This very real possibility is suggested, moreover, by feminist scholarship that has overtaken the "biological sex" versus "social gender" distinction outlined in Rubin (1975) that inspired Shiao et al. (2012). These more recent examinations of science and the construction of difference offer a different lesson than the one taken by the authors, namely, that our very best, most earnest approximations of the natural world are always shot through with the social (Harding 1986; E. Martin 2001; Schiebinger 1993). Although the "social race" versus "biological clinal class" binary might be of theoretical use in the abstract, the fundamentally social nature of both our perceptions and our measurements makes it extremely difficult to gauge exactly how much if at all our racial categories correspond to any independent reality of clusters, classes, or clines.

In short, convergences between clinal class and race are not only likely to be the exception rather than the rule, but further, the occasional alignments between the two can be explained by two thoroughly social processes of construction. On one hand are the ways in which the social forces of categorization, segregation, and stratification can create endogamous racial or ethnic groups that acquire over time the genetic distinctiveness that later leads to their measurement as biological clinal classes. On the other is the way in which, as numerous social studies of sciences have shown (Abend 2006; Bloor 1991; Fleck [1935] 1979; Shapin 1995), widely shared beliefs about the world make their way into both the practice of science and the knowledge it produces (with respect especially to race and genetics, see Bolnick et al. 2007; Fujimura and Rajagopalan 2011; Fullwiley 2007; Nelkin and Lindee 1995; Wailoo 2001; Wailoo and Pemberton 2006). As a result, social theories and scientific measurements can acquire a "self-vindicating" quality (Hacking 1992). It is to this second possibility that I now turn, exploring empirically how DNA-based human population clusters are statistically constructed and where human manufacture and cultural belief can "get in" to the numbers.

THE SOCIAL CONSTRUCTION OF GENETIC POPULATION CLUSTERS

When Shiao et al. (2012) write that "recent research on the human genome challenges the basic assumption that human races have no biological basis" (p. 68), they are largely referring to a mid-2000s literature on population structure that appeared in leading journals, such as *Genetics* and the *American Journal of Human Genetics*. From this body of work, they deduce that "it is now possible to statistically distinguish human subspecies that overlap on many biological characteristics" (Shiao et al. 2012:70) and that these "statistically identified genetic clusters appear to be concordant with…racial/ethnic self-identification" (Shiao et al. 2012:71). In the absence of empirical scholarship on "clinal classes," this is the literature that provides the basis for Shiao et al.'s theory about the "biological reality" of race.

To address the arguments that Shiao et al. (2012) make about patterns of human biological variation, I review the genetic research reports on which they rely: Pritchard, Stephens, and Donnelly (2000); Wilson et al. (2001); Risch et al. (2002); Rosenberg et al. (2002, 2005); and Paschou et al. (2007). In addition, I examine Burchard et al. (2003), written by several of the same authors as Risch et al. (2002), and Serre and Pääbo (2004), which posed important questions about how research design influences estimates of human population structure. It is worth noting that even though Shiao et al. focused on genetic research published in the early and mid-2000s, scientific inquiry into human population structure has

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

continued apace, so the ideas raised here are still relevant to the current production of knowledge about genetic clusters, particularly as they relate to racial and ethnic admixture (see, for example, Hellenthal et al. 2014; Heyer et al. 2009; Shah et al. 2011).

The Statistical Inference of Genetic Population Clusters

The genetics literature upon which Shiao et al. (2012) draw reports on researchers' efforts to develop and/or use statistical methods with genetic data to identify subgroups or "clusters" in the human species and to assign individuals to these groupings. As Shiao et al. note,

> [t]he primary tools for identifying population structure have been, in order of emergence, (1) comparisons of predefined populations, (2) the Bayesian clustering approach of the program [*Structure*] developed by Pritchard, Stephens, and Donnelly (2000), and (3) new variations on the classical technique of principal components analysis (Paschou et al. 2007). (Shiao et al. 2012:72)

The first of these has been called the "ancestry-informative markers" (AIMS) approach, because it uses genetic loci (markers) that have been selected precisely due to the fact that they vary more than others between the populations that researchers have predefined as distinct clusters or ancestry groups. In the ancestry-informative approach, the analyst simply chooses in advance the ancestry groups that he or she believes exist—for example, European, African, Asian—and then searches for genes whose alleles (i.e., variants) differ most sharply in their frequency between these predetermined clusters. This is clearly a highly subjective process, which can—and often does—involve using off-the-shelf social categories. Pritchard et al. (2000:945) described these methods as "typically subjective, based, for example, on linguistic, cultural, or physical characters, as well as the geographic location of sampled individuals. . . . [I]t may be difficult to know whether a given assignment of individuals to populations based on these subjective criteria represents a natural assignment in genetic terms."

In contrast, both the Bayesian *Structure* program and principal components analysis (PCA) do not presume which or how many clusters or ancestry groups exist but rather seek to infer these. *Structure*'s principal output is *K*, the inferred number of clusters, as well as the assignment of individuals to these estimated clusters. PCA focuses on identifying the single nucleotide polymorphisms (SNPs)—that is, the "single letter variations" in DNA strands composed of nucleotides labeled *A, C, G*, and *T*—that most efficiently convey the structure of variation within the entire sample; Paschou et al. (2007) call these "structure informative" or "PCA-correlated" SNPs, and they argue that these selected markers are better able to delineate clusters than randomly selected ones would.

As this line of research has developed, it has become clear that the subgroups so identified are statistical artifacts, shaped by researchers' techniques and assumptions. This should come as no surprise; as the sociology of scientific knowledge has repeatedly demonstrated, the same could be said of any science (see, for example, Fleck [1935] 1979; Schiebinger 1993). Nonetheless, it is important to keep in mind that although genetic researchers may write as if they have simply "discovered," "revealed," or "discerned" subpopulations that objectively exist in the world independently of human agency, in fact these are collectives they have constructed. Serre and Pääbo (2004) make this abundantly clear when they refer to the product of such calculations, including their own, as "inferred populations."

The socially constructed nature of our understandings of human biological variation today becomes most apparent when we examine the assumptions and techniques that go into

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

them. The sampling of individuals, the genetic information analyzed, and the level of resolution desired as well as assumptions about human evolutionary history, "major" populations, and admixture all have an impact on which clusters—and how many—statistical analyses produce. In other words, in addition to complex statistical methods, powerful computers, and dizzying amounts of genetic data, researchers also bring presumptions and expectations to their work, and all of these contribute to the portrait of the human species that results. Below, I outline key aspects of research design that statisticians and geneticists have identified as influencing their estimates of human population structure.

Sample Size. It comes as no surprise that the number of individuals whose genotypes are analyzed will have an impact on the number and content of inferred clusters. As with statistical analysis more generally, larger samples permit greater confidence in claiming significant distinctions between groups. Pritchard et al. (2000:955) note moreover that "the accuracy of the assignments [of individuals to clusters] depends on a number of factors, including the number of individuals." And within a sample, the relative shares contributed by different groups also influence the estimation of clusters. As Rosenberg et al. (2002:2384) put it, "[g]roups with larger sample size are also more easily separated" into distinct clusters (see also Rosenberg et al. 2005; Serre and Pääbo 2004).

Sample Geography. Similarly, it is not hard to understand why the geographic location of the sampled individuals makes a difference. To take an extreme example, analysis of a data set that included only people from either Colombo, Sri Lanka, or Reykjavik, Iceland would likely conclude that our species is made up of two clusters. And given the great distance between the two places, the analysis would probably assign the individuals sampled to one cluster or the other with ease, since genetic distinctiveness is positively correlated with geographic distance (Templeton 1999). But few people would be satisfied that sampling from just two locations, Colombo or Reykjavik, would give us a good picture of the structure of the entire human species around the globe. In other words, the representativeness and the geographic source locations of the sample shape our conclusions about population structure.

This problem is not as remote as the hypothetical example above suggests. Because we do not (yet) have vast data banks of genetic data collected in every corner of the world (at least not at the disposition of scientific researchers), claims about genetic clusters of human beings—that is, about the structure found within a species of more than 7 billion people—are often based on samples that are relatively very small. Pritchard et al. (2000) tested their *Structure* algorithm with a sample of 72 Africans and 90 Europeans; Rosenberg et al. (2002) published an article in *Science* entitled "Genetic Structure of Human Populations" based on a sample of 1,056 individuals; and Paschou et al. (2007) sought to demonstrate the utility of their PCA method with a sample of 274 people. Moreover, large "sampling gaps" in the data available clearly skew the picture of human genetic diversity (Serre and Pääbo 2004:1682; see also Wilson et al. 2001:268 on the need for "geographically exhaustive" data). When Serre and Pääbo (2004) analyzed the widely used HGDP-CEPH Human Genome Diversity Cell Line Panel,[2] they found not only a dearth of individuals from North Africa, for example, but a complete absence of indigenous people from North America.

Given the relatively small numbers and limited locations of human beings who have been genotyped, the distribution of individuals sampled is important for any assessment of population structure. Serre and Pääbo (2004) argued that sampling often concentrates on "the extremes of continental land masses" (p. 1680), maximizing the geographic and therefore genetic distance between individuals presumed to belong to distinct continental clusters. Without "a sampling strategy that maximizes the geographic distribution of samples and keeps similar sample size for each geographical area," they warned, researchers risked

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

falsely creating "apparent substructures" (Serre and Pääbo 2004:1681). In contrast, when these researchers designed a study that sampled individuals "such that their geographic distribution around the world approximates the distribution of the human population as a whole and includes areas where Africa, Asia, and Europe meet," the pattern of genetic variation they found was "one of gradients of allele frequencies that extend over the entire world, rather than discrete clusters" (Serre and Pääbo 2004:1679-1680). Not only did their homogenous sampling strategy paint a different picture of clustering in the human species than previous samples had, but it also yielded much higher estimates of individuals' admixture. They concluded that "the discrete clusters described by Rosenberg et al. (2002) . . . might be caused by discontinuities in the sampling, because when samples that have equal numbers of individuals of each population are analyzed. . . . The inferred populations yielded by *Structure* do not match continents or geographical regions but represent theoretical 'populations' in which all individuals show admixture to at least two such 'populations'" (Serre and Pääbo 2004:1682). Although in their subsequent analysis of study design impact on clustering estimates, Rosenberg et al. (2005) argued that a sample's geographic dispersion had only small and inconsistent effects on clustering, their original study concluded that "[b]ecause sampling was population based, the sample likely produced clusters that were more distinct than would have been found in a sample with random worldwide representation" (Rosenberg et al. 2002:2384).

Genetic Data Collected. As with the individual human beings sampled for analyses of population structure, the number and types of genetic markers studied also have repercussions for geneticists' conclusions about clustering (Rosenberg et al. 2005; Wilson et al. 2001). Pritchard et al. (2000) found that with fewer loci (locations in the genome) typed, it was harder both to determine the number of clusters in the data and to accurately assign people to the inferred clusters.

Different types of genetic data can be collected from a person's DNA, and the kind an analyst chooses has implications for the inference of population clusters. Wilson et al. (2001) compared results obtained from using (1) chromosome 1 microsatellites and (2) X-linked microsatellites. With the first type of genetic data, they discerned (K =) four clusters among human beings, but when they turned to the second, they inferred three. Both the AIMs and PCA approaches are premised on the idea that certain genetic markers are better than others when it comes to describing clusters. AIMs, however, are tautological constructs; they have been cherry-picked precisely for their power to yield the groupings that their users have already decided are clusters. And although PCA-correlated SNPs do not rely on the predetermination of clusters, their selection is nonetheless limited by the data from which they are derived. As Paschou et al. (2007) put it, "SNPs chosen for ancestry inference in one continent are in general no better than random SNPs and not transferable to other continents" (p. 1678). A corollary of this finding is that the characteristics of samples described above that influence structure inference—that is, number and geographic location of individuals sampled as well as number and type of loci—also shape PCA output. In fact, "each time a new population is added to the analysis, the panel of SNPs needed for population differentiation is modified" (Paschou et al. 2007:1683). Although Paschou et al. claim that their nonparametric approach "does not rely on any assumptions or modeling of the data" (Paschou et al. 2007:1681), they gloss over the fact that the very data themselves embody certain presumptions about human variation, due to the ways in which they were selected. The distillation of such data into a few PCA-correlated SNPs crystallizes prior sampling decisions into a particular vision of population structure.

Level of Analytical Resolution. A common thread running through analyses of human population structure is the observation that with more data—more people, in more locations, typed

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

at more genetic loci—researchers can discern ever-finer distinctions between groups. With more data, they can drill down to lower levels of analysis or, to put it differently, can work at higher levels of resolution. The idea is that fewer genetic data are needed to distinguish, say, Europeans from Native Americans than are needed to distinguish Spaniards from Greeks (Leroi 2005). This raises the question of which level of resolution if any is the "natural," "best," or "preferable" one for identifying human clusters. Although several researchers have suggested that different levels of resolution might be appropriate for different kinds of inquiry—for example, in relation to drug response (Wilson et al. 2001), local evolutionary history (Serre and Pääbo 2004), or regional population structure (Paschou et al. 2007)—the analyst's level of resolution, or geographical focus, is rarely treated explicitly as a variable influencing outcomes. Yet "the complexity of human demographic history means that there is no obvious natural clustering scheme, nor an obvious appropriate degree of resolution" (Wilson et al. 2001:265). Accordingly, Shiao et al. (2012:73) insist that their clinal classes are "scalable," meaning that "the concept of clinal class does not require a fixed number of classifications, as the number of classes depends on the attempted resolution for detecting population structure."

In the genetics literature, a more common way to talk about the relationship between data samples and the level of cluster resolution is to describe the amount of data "needed" to discern certain human groups that the researcher has already determined should be distinguished. For example, Rosenberg et al. (2005) report needing many more loci in order to detect internal structure in the Middle East, and they conclude that "if *enough* markers are used with a *sufficiently* large worldwide sample, individuals can be partitioned into genetic clusters that match major geographic subdivisions of the globe" (p. 660; emphasis added). In both cases, the inquiry begins with the assumption that there are meaningful genetic clusters to be detected—"internal structure" in the Middle East, or "major geographic subdivisions of the globe"—and so the analyst's task is simply to assemble the quantity of data required to "find" them. Little mention is made of the ways in which these are exercises in using data to "reveal" the divisions that one has set out to demonstrate.

Even less effort goes into pointing out the ways that the divisions we seek to corroborate originate in socially—as opposed to genetically—relevant taxonomies. Risch et al.'s (2002:7) comment on the demarcation of whites and Latinos in the United States offers an illuminating example: "As expected, differentiating Caucasians and Hispanic Americans, who are admixed but mostly of Caucasian ancestry, is more difficult and *requires* a few hundred random STRPs [short tandem repeat polymorphisms] or about 50 highly selected loci" (emphasis added). It is no accident that Risch et al. find distinguishing "Caucasians" from "Hispanic Americans" a meaningful exercise despite their genetic relatedness; the boundary between them is highly socially salient in the United States today. Although the language here is of natural, objectively significant genetic difference that simply "requires" more data to be registered, at issue is not the mapping but the construction of human clusters.

Human Evolutionary History. Fundamental to any search for genetic clusters in the human species is the presumption that such groupings exist. But this idea is debated. As some argue, clines—that is, continuous gradients of biological variation—may be a better way to think about genetic patterns among humans. When Wilson et al. (2001:267) had difficulty identifying stable clusters in their data, they concluded that it was "probably because there are no natural clusters, as there has not been a history of bifurcation in human populations." Similarly, Serre and Pääbo (2004:1679) found that "when individuals are sampled homogeneously from around the globe, the pattern seen is one of gradients of allele frequencies that extend over the entire world, rather than discrete clusters. Therefore, there is no reason to assume that major genetic discontinuities exist between different continents or 'races.'"

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

Another consequential assumption about human evolutionary history to influence estimates of population structure has to do with whether allele frequencies across groups should be modeled as correlated or uncorrelated with each other. The rationale for the former is that "[t]he correlated frequencies model . . . supposes that the various clusters represent populations that have descended with genetic drift from a common ancestral population, so that alleles in different clusters have correlated frequencies due to shared ancestry" (Rosenberg et al. 2005:661). In contrast, Serre and Pääbo (2004) favor the uncorrelated model because it "would best represent a situation in which colonizations of various parts of the world originated from ancestral populations in which genetic drift would have been strong enough to allow microsatellite allele frequencies to become independent from each other" (p. 1681). The choice of correlation model appears to be related to the number of clusters that the program *Structure* infers. Pritchard et al. (2000:955-956) contend that allowing allele frequencies to be correlated across populations may lead researchers to estimate larger numbers of clusters, while the uncorrelated model may risk "merg[ing] subpopulations that share similar frequencies"—that is, underestimating the number of clusters. In other words, researchers' presumptions about the origins of the current landscape of human genetic variation are another factor that shapes eventual conclusions about clusters.[3]

Number of Clusters: What Are the "Major" Populations of Interest? One of the most powerful ways in which geneticists' racial preconceptions can shape their analyses of human population structure lies in their assumptions about which clusters and/or how many characterize our species. In the AIMs approach, the analyst simply dictates the clusters he or she feels exist, independent of any genetic data. The developers of the *Structure* and PCA methods argue that their approaches are less subjective, because they permit the analyst to estimate the number of clusters found in the genetic data rather than simply invent a number. Yet even these methods turn out to offer several entry points for researchers' preconceptions to influence the analysis. For one thing, it is common for scientists to test the model fit of only a few possible options for the number of clusters (K), ranging in the single digits (e.g., $K = 2$ through 6 in Rosenberg et al. 2005), which is consistent with contemporary notions of the number of races in our species; the U.S. federal classification system, for example, includes five races at present (i.e., white, black, American Indian or Alaska Native, Asian, Native Hawaiian or other Pacific Islander; see U.S. Office of Management and Budget 1997). Paschou et al. (2007) manually set the number of clusters to four, although they base their choice on estimates of how much more an additional component would explain (p. 1675).

Another potential entry point for researchers' subjective assumptions is through the assessment of which methods seem to produce "sensible" results, or outcomes that corroborate the expected population categories. In tackling the "notoriously difficult" problem of estimating the number of clusters (K) in a sample, Pritchard et al. (2000:949) note that their Bayesian approach calls for determining a posterior distribution of K that is "peculiarly dependent on the modeling assumptions made." Consequently, they propose and use an "*ad hoc* and computationally convenient" strategy, but they admit that its assumptions "are dubious at best, and we do not claim (or believe) that our procedure provides a quantitatively accurate estimate of the posterior distribution of K." Its "main justification," the authors go on to report, is that "it seems to give sensible answers in practice." And as we have seen, what many geneticists seem to find sensible is a small number K of genetic clusters that corresponds to lay beliefs about the number of races.

Similarly, the efficacy of clustering methods is usually tested or demonstrated by comparing their results to "known" population or racial classifications. Pritchard et al. (2000) illustrate the virtues of *Structure* by showing how it differentiates groups that "should" be

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

differentiated: Africans and Europeans. As noted previously, Risch et al. (2002:7) promote structure inference methods that successfully distinguish European and Hispanic Americans. Their conviction that "Caucasians" and "Hispanic Americans" are distinct clusters that should be distinguished—even though the latter are in large part recent descendants of the former—is telling. The classificatory distinction between them is one that has arisen and gained meaning in the United States due to social and political circumstances; the label *Hispanic* is only about 50 years old, as is the concept that people from diverse corners of the former Spanish empire—like Peruvians, Mexicans, and Dominicans—somehow naturally belong together (Graham 2002). So when Risch et al. (2002) tout the power of methods that can distinguish "Hispanics" (who are not actually included as a race in the current U.S. federal classification scheme) from "Caucasians," they are calibrating a statistical method to social conventions, not to an objective genetic landscape.

Finally, scientists' preconceptions about which clusters characterize our species also influence their interpretations of the number *K* of groupings that their algorithms produce. In a striking example, Pritchard et al.'s (2000) analysis suggests that their genotype data from Africans and Europeans would be better described as forming three, four, or five clusters than merely two. In other words, a racial "black/white" binary structure is not supported by their method. Yet they go on to conclude that "our methods find it quite easy to separate the two continental groups into the correct clusters" (Pritchard et al. 2000:952), reverting to a racial binary that is not indicated by their statistical analysis. In addition, the authors point out elsewhere that interpreting the meaning of the inferred value of *K* is far from straightforward, giving two hypothetical illustrations. One comes from their simulation of an admixed population whose structure they suggest could be understood in two ways. One would be to characterize it as having two clusters, assigning individuals to either of the two ancestral populations modeled in the exercise. Another interpretation, however, would be to view the population as having five clusters, grouping individuals by the number of grandparents they have from one of the ancestral populations (i.e., according to whether they had zero, one, two, three, or all four grandparents from one of the two ancestral populations). The other hypothetical example they use to highlight the challenges of making sense of an inferred number of clusters *K* is the following:

> [C]lusters may not necessarily correspond to "real" populations. . . . [I]magine a species that lives on a continuous plane, but has low dispersal rates, so that allele frequencies vary continuously across the plane. If we sample at K distinct locations, we might infer the presence of K clusters, but the inferred number K is not *biologically* interesting, as it was determined purely by the sampling scheme. All that can usefully be said in such a situation is that the migration rates between the sampling locations are not high enough to make the population act as a single unstructured population. (Pritchard et al. 2000:956)

In this case, it is not clear whether to speak of an unstructured population (i.e., a single population without internal clusters) or one with two or more clusters. Pritchard et al.'s warning that clusters may not be the same thing as "'real' populations" echoes Serre and Pääbo's (2004) suggestion that inferred clusters may best be thought of as "theoretical populations."

The assumptions that geneticists make about major human clusters have a corollary in their presumptions about which peoples are admixed. The *Structure* program allows researchers to incorporate estimates for how admixed a sample may be; specifically, the analyst may "specify a distribution for *Q* [admixture proportions for each individual], which

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

in general will depend on the type and amount of admixture we *expect* to see" (Pritchard et al. 2000:948, emphasis added). Again, preconceptions—in this case, about admixture—are integrated into the method. In an unsubstantiated nod to the U.S. one-drop rule, Burchard et al. (2003:1173) claim that "[d]espite the admixture [with whites], black Americans, as a group, are still genetically similar to Africans." Consistent with centuries of American "one-drop" ideology, African Americans—despite their non-African ancestry—can simply be understood as black.

The above excursion into the literature on human genetic clusters from which Shiao et al. (2012) draw has aimed to take a close look at the empirical evidence for their claim that contemporary genetics has identified recognizable groupings in our species. What emerges from the examination of diverse methods that Shiao et al. tout as demonstrating the objective existence of human clusters is that there is clearly a great deal of room for analysts' assumptions about human biological variation to inform every stage of research design and thus to produce a particular outcome. As described above, choices about the data sampled, the specification of models, beliefs about the course of human evolutionary history, the number and kind of clusters set, and the interpretation of results all color geneticists' depictions of human population structure. This is not to belittle the work of scientists who have thought deeply about biological heterogeneity and its origins in our species, and who are the first to note some of the implications of their research designs. It is, however, to argue that genetics, like any other science, bears the imprint of the society in which it is cultivated and that our methods of building knowledge necessarily incorporate our subjective perspectives. The many turns at which such presumptions can shape our conclusions about the existence of genetic human clusters recalls Ian Hacking's (1992) idea of "the self-vindication of the laboratory sciences," in which we can mutually adjust theories, data, instruments, and analyses—"ideas," "things," and "marks," he calls them—to arrive at our desired or anticipated conclusions. It is also reminiscent of what we know about the conduct of "race science" in the nineteenth through twentieth centuries (Dubow 1995; Gould 1996; Nobles 2000; Smedley and Smedley 2012; Stepan 1982).

ARE GENETIC CLUSTERS RACES?

As quoted previously, Shiao et al. (2012) make the major empirical claim that "statistically identified genetic clusters appear to be concordant with . . . racial/ethnic self-identification" (p. 71). To examine the evidence for this argument, I turn once again to the same body of genetic research they cite. Strikingly, I find that for the most part, these scientists do *not* link human population structure to race, and in some cases, they explicitly argue *against* the interpretation of their findings as supporting the race concept. (For similar findings from interviews with geneticists, see Bliss 2012; Morning 2011.) In the mid-2000s literature on the estimation of human population structure to which Shiao et al. point, only two authors (Neil Risch and Hua Tang) seem to have consistently equated statistical clusters with "races."

Some of the scientific commentary on population structure foregoes the concept of race altogether. Pritchard et al. (2000), the developers of *Structure*, do not use the term, and their results with African and European data do not map on to a binary black/white racial distinction; instead, they find "population structure within the continental groupings." Nor does their method "differentiate . . . between the Europeans and Asians with great accuracy using this data set" (Pritchard et al. 2000:952). Similarly, Paschou et al. (2007), proponents of PCA for the estimation of population structure, make no mention of "race" but only of populations and "subpopulations." Rosenberg et al.'s (2002) study also eschews the term *race* in favor of "self-reported population ancestry," where *population* refers to much finer-grained

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

groupings than races, such as "Daur, Hezhen, Mongola, Oroquen, and Xibo, all from northern China" or "Balochi, Makrani, Pathan and Sindhi" (p. 2383). Nor do their clusters map onto traditional races; they infer six main genetic clusters, of which one corresponds to the Kalash of Pakistan; in a subsequent 2005 analysis, their sixth cluster varies between splitting off part of the Native American population or identifying the Kalash (Rosenberg et al. 2005:662).

At other times, researchers argue strenuously against the imposition of social racial categories onto the map of human genetic variation. Wilson et al. (2001:266) contend that "commonly used ethnic labels (such as Blacks, Caucasian and Asian) are insufficient and inaccurate descriptions of human genetic structure." For one thing, their analysis supported a partition of humans into four clusters that corresponded roughly to "Western Eurasia," "Sub-Saharan Africa," "China," and "New Guinea" but that included populations that would have been classified differently by race. Sixty-two percent of the Ethiopians in the sample were classified in the first cluster, along with most Jews, Norwegians, and Armenians ("indicating that placement of these individuals in a 'Black' cluster would be an inaccurate reflection of the genetic structure"; Wilson et al. 2001:266), and only 24 percent were placed in the cluster with other ostensible "blacks" like the Bantu. Twenty-one percent of the Afro-Caribbeans were assigned to the West Eurasian cluster. And populations from China and New Guinea ended up in separate clusters, "indicating that the ethnic label 'Asian' is also an inaccurate description of population structure" (Wilson et al. 2001:267). Moreover, Wilson et al. found that when comparing racial groups to inferred genetic clusters, the former never did a better job of capturing patterns of drug-metabolizing enzyme allele frequencies than genetic clusters did. Rosenberg et al. (2005:668) make a point of writing, "Our evidence for clustering should not be taken as evidence of our support for any particular concept of 'biological race.'" In their view, concepts like clines and clusters are scientifically useful because they "facilitate further research into topics such as human evolutionary history and the identification of medically important genotypes that vary in frequency across populations" (Rosenberg et al. 2005:668); in contrast, "[t]he arguments about the existence or non-existence of 'biological races' in the absence of a specific context are largely orthogonal to the question of scientific utility" (Rosenberg et al. 2005:669). Finally, Serre and Pääbo (2004:1679) argue there is "no reason to assume that major genetic discontinuities exist between different continents or 'races.'"

Overall, the empirical literature in human genetics that Shiao et al. (2012) report as having demonstrated the existence of biologically demarcated races is far from doing so. Indeed, of the eight articles reviewed here, only two—Risch et al. (2002) and a subsequent article by the same authors along with additional coauthors (Burchard et al. 2003)—make the claim that races correspond to inferred genetic clusters. (When Ziv and Burchard coauthored with Paschou et al. [2007], they dropped mention of race altogether.) These two publications are also notable because they are the only ones reviewed here that are not based on original research; they are surveys of existing scholarship. When Shiao et al. claim then that "the increasing power of computer assisted quantitative data analysis and the growing resolution of available genetic data has enable quantitative geneticists to identify an empirical structure within human genetic variation that at a certain scale resembles the continentally based racial classifications of the U.S. federal government" (Shiao et al. 2012:68), only one of the five supporting citations they provide is a report of original research findings.

Why do Risch et al. (2002) and Burchard et al. (2003) put forth a "genetic cluster as race" argument that is at odds with the other contemporary writings on human population structure studied here? Examination of their publications reveals that both redraw the boundaries of racial groups in ways that conform to inferred genetic clusters. Although Risch et al.'s

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

(2002:3) claim that contemporary genetic research corroborates "the classical definition of races based on continental ancestry—namely African, Caucasian (Europe and Middle East), Asian, Pacific Islander . . . and Native American" is meant to suggest that today's science proves the accuracy of taxonomies developed in the 1700s, scrutiny of their article suggests that the match is due instead to the lingering influence of eighteenth-century conjectures on scientific analyses in the twenty-first. For example, Risch et al. support the biological reality of a "Caucasian" category by pointing to geneticists' inference of a cluster that includes South Asians. This expanded mapping of the Caucasian population runs counter not only to contemporary racial categories in the United States, Canada, and the United Kingdom (Morning 2008; U.S. Office of Management and Budget 1997) but also to traditional understandings of whiteness (Haney López 1996; Morning 2001). Similarly, Risch et al. include North Africans and Middle Easterners in their Eurasian cluster, which, although it accords with current U.S. federal racial categorization, also conflicts with popular notions of whiteness in the United States and western Europe and possibly with the racial categories that will appear on the 2020 U.S. Census (Compton et al. 2012; Kayyali 2013; Samhan 1999). In other words, it is easy to say that genetic studies prove the existence of race if you change the content of your racial categories to match the genetic ones. And the disparity that gets erased by such sleight of hand is an enormous one: Burchard et al. (2003:1172) identify the assignment of "South, Central, and West Asians" as the primary difference between census and genetic categorization, which means roughly 2 billion people out of the over 7 billion alive today. Scientific analyses that classify over a quarter of the world's population differently than racial taxonomies hardly furnish evidence of a match between estimates of genetic variation on one hand and racial categories on the other.

The previous pages have assessed at length the empirical heart of Shiao et al.'s (2012) argument that the science of human genetics has identified biological clusters within our species that correspond to what Americans (and others) popularly call "races." Examining the mid-2000s literature on human population structure from which Shiao et al. draw inspiration, however, I come to the opposite conclusion. First, although it is true that geneticists have sought to infer clusters within the global population, the statistical groupings that result are not so much "natural," objective subpopulations that scientists simply "discover" as they are collectives that analysts construct. As their makers readily admit, the number and content of such clusters depend on a variety of assumptions, including those that contribute to the shaping of the genetic data sets used. Second, few participants in scientific debate about population structure seem to find "race" a useful analytical tool, let alone equate it with statistically derived, DNA-based clusters.

CONCLUSION

Upon inspection of both Shiao et al.'s (2012) theory of clinal class and the research literature in human genetics that supports it, contemporary genomic research seems to pose little challenge to the theory of race as socially constructed. Not only can constructivist theory accommodate or explain the occasional alignment of social classifications and genetic estimates that Shiao et al.'s model hypothesizes, but empirical research on human genetics is far from claiming—let alone demonstrating—that statistically inferred clusters are the equivalent of races.

The authors nonetheless render the scholarly community the important service of challenging sociologists to examine the knowledge produced by contemporary biologists, consider how it meshes with their own, and attempt to account for any divergences. Interdisciplinary comparison prods us to refine, clarify, and perhaps better disseminate our

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

theories, very much in the spirit of Shiao et al.'s (2012) article. Such efforts are relevant not only to scholars of race, moreover; the role of biology is debated with respect to countless societal outcomes (Nelkin and Lindee 1995). More broadly, reflection on the implications of biological findings and claims has a special urgency in a time when, as anthropologist Nina Jablonski put it, "genomic knowledge widens" (Messer 2014).

Finally, I wish to concur with Shiao et al.'s (2012) call for "the removal of obsolete claims about human biology from racial constructionism" (p. 69). The field to be scrutinized, however, should be expanded from constructivism to include all the sciences, natural and social. The general idea that the human species is characterized by biologically distinguishable races is by now an old one, going back several centuries in the West, and there is no question that it contains an accumulation of beliefs about human difference that not only predate the contemporary sciences but run counter to how we think today about human society and biology. Mapping social races to genetically inferred clusters, moreover, is the latest version of a centuries-old effort to identify a biological basis for racial groupings. Yet perhaps paradoxically, this effort speaks to the ineffably social nature of race. Its power and centrality derive from its social meaningfulness, and the attempt to ground it in genetic science reveals the enduring importance of enlisting our most authoritative bodies of knowledge to verify its objective reality.

ACKNOWLEDGMENTS

The author wishes to thank Catherine Lee, Alondra Nelson, and Wendy Roth for their advice and comments.

NOTES

1. Note the role of subjective perceptions of distinctiveness in racial categorization, consistent with Weber's ([1956] 1978) insights about the importance of belief in ethnic classification. Further exploration of the interplay between belief and sensory perception can be found in Osagie Obasogie's (2014) *Blinded by Sight: Seeing Race through the Eyes of the Blind*, which challenges the commonplace notion that we know race when (or because) we passively "just see" it.
2. See http://www.cephb.fr/en/publis_HGDP.php for a list of publications based on this panel.
3. Other statistical assumptions that shape the inference of population structure include whether "marker loci are unlinked and at linkage equilibrium with another within populations" (Pritchard, Stephens and Donnelly 2000:946), Hardy-Weinberg equilibrium characterizes populations, migration is frequent or infrequent, geographic location is a good proxy for ancestry, and which particular statistical distribution certain variables follow (Pritchard et al. 2000).

REFERENCES

Abend, Gabriel. 2006. "Styles of Sociological Thought: Sociologies, Epistemologies, and the Mexican and U.S. Quests for Truth." *Sociological Theory* 24(1):1-41.

American Anthropological Association. 1998. "Statement on 'Race.'" *American Anthropologist*. Retrieved August 19, 2014 (http://www.aaanet.org/stmts/racepp.htm).

American Sociological Association. 2003. *The Importance of Collecting Data and Doing Social Scientific Research on Race*. Washington, DC: American Sociological Association.

Baer, Roberta D., Erika Arteaga, Karen Dyer, Aimee Eden, Rosalyn Gross, Hannah Helmy, Margaret Karnyski, Airia Papadopoulos, and Doug Reeser. 2013. "Concepts of Race and Ethnicity among Health Researchers: Patterns and Implications." *Ethnicity and Health* 18(2):211-25.

Bliss, Catherine. 2012. *Race Decoded: The Genomic Fight for Social Justice*. Redwood City, CA: Stanford University Press.

Bloor, David. 1991. *Knowledge and Social Imagery*. Chicago: University of Chicago Press.

Bolnick, Deborah A., Duana Fullwiley, Troy Duster, Richard S. Cooper, Joan H. Fujimura, Jonathan Kahn, Jay Kaufman, Jonathan Marks, Ann Morning, Alondra Nelson, Pilar Ossorio, Jenny Reardon, Susan

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

M. Reverby, and Kimberly TallBear. 2007. "The Science and Business of Genetic Ancestry." *Science* 318(5849):399-400.

Burchard, Esteban Gonzalez, Elad Ziv, Natasha Coyle, Scarlett Lin Gomez, Hua Tang, Andrew J. Karter, Joanna L. Mountain, Eliseo J. Perez-Stable, Dean Sheppard, and Neil Risch. 2003. "The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice." *New England Journal of Medicine* 348(12):1170-75.

Calhoun, Craig, Donald Light, and Suzanne Keller. 2001. *Understanding Sociology*. New York: Glencoe McGraw-Hill.

Compton, Elizabeth, Michael Bentley, Sharon Ennis, and Sonya Rastogi. 2012. "2010 Census Race and Hispanic Origin Alternative Questionnaire Experiment," edited by 2010 Census Program for Evaluations and Experiments. Washington, DC: U.S. Census Bureau.

Danna, Karen. 2011. "The Organization of Everyday Identity (and What It Reveals about the Organization of Culture)." Paper presented at American Sociological Association, August 20, Las Vegas, NV.

Dubow, Saul. 1995. *Scientific Racism in Modern South Africa*. Cambridge, UK: Cambridge University Press.

Duster, Troy. 2006. "American Sociological Association 2005 Presidential Address: Comparative Perspectives and Competing Explanations. Taking on the Newly Configured Reductionist Challenge to Sociology." *American Sociological Review* 71(1):1-15.

Fleck, Ludwik. [1935] 1979. *Genesis and Development of a Scientific Fact*. Chicago: University of Chicago Press.

Fujimura, Joan H. and Ramya Rajagopalan. 2011. "Different Differences: The Use of 'Genetic Ancestry' versus Race in Biomedical Human Genetic Research." *Social Studies of Science* 41(1):5-30.

Fullwiley, Duana. 2007. "Race and Genetics: Attempts to Define the Relationship." *Biosocieties* 2(2):221-37.

Gould, Stephen Jay. 1996. *The Mismeasure of Man*. New York: Norton.

Graham, Hugh Davis. 2002. "The Origins of Official Minority Designation." Pp. 288-99 in *The New Race Question: How the Census Counts Multiracial Individuals*, edited by J. Perlmann and M. C. Waters. New York: Russell Sage Foundation and the Levy Economics Institute of Bard College.

Hacking, Ian. 1992. "The Self-vindication of the Laboratory Sciences." Pp. 29-64 in *Science as Practice and Culture*, edited by A. Pickering. Chicago: University of Chicago Press.

Haney López, Ian F. 1996. *White by Law: The Legal Construction of Race*. New York: New York University Press.

Harding, Sandra. 1986. *The Science Question in Feminism*. Ithaca, NY: Cornell University Press.

Harris, Marvin. 1964. *Patterns of Race in the Americas*. Westport, CT: Greenwood Press.

Hellenthal, Garrett, George B. J. Busby, Gavin Band, James F. Wilson, Cristian Capelli, Daniel Falush, and Simon Myers. 2014. "A Genetic Atlas of Human Admixture History." *Science* 343:747-51.

Helweg, Arthur and Usha Helweg. 1990. *An Immigrant Success Story: East Indians in America*. Philadelphia: University of Pennsylvania Press.

Heyer, Evelyne, Patricia Balaresque, Mark A. Jobling, Lluis Quintana-Murci, Raphaelle Chaix, Laure Segurel, Almaz Aldashev, and Tanya Hegay. 2009. "Genetic Diversity and the Emergence of Ethnic Groups in Central Asia." *BioMed Central Genetics* 10(1):49.

Humes, Karen R., Nicholas A. Jones, and Roberto R. Ramirez. 2011. *Overview of Race and Hispanic Origin: 2010*. Washington, DC: U.S. Census Bureau.

Jacobson, Matthew Frye. 1998. *Whiteness of a Different Color: European Immigrants and the Alchemy of Race*. Cambridge, MA: Harvard University Press.

Kayyali, Randa. 2013. "US Census Classifications and Arab Americans: Contestations and Definitions of Identity Markers." *Journal of Ethnic and Migration Studies* 39(8):1299-1318.

Kevles, Daniel J. 1985. *In the Name of Eugenics: Genetics and the Uses of Human Heredity*. Cambridge, MA: Harvard University Press.

Krieger, Nancy. 2005. "Stormy Weather: *Race*, Gene Expression, and the Science of Health Disparities." *American Journal of Public Health* 95(12):2155-60.

Lee, Catherine. 2009. "'Race' and 'Ethnicity' in Biomedical Research: How Do Scientists Construct and Explain Differences in Health?" *Social Science & Medicine* 68:1183-90.

Leroi, Armand Marie. 2005. "A Family Tree in Every Gene." *The New York Times*, March 14, p. A21.

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

Marks, Jonathan. 1995. *Human Biodiversity: Genes, Race, and History*. New York: Aldine de Gruyter.
Martin, Emily. 2001. *The Woman in the Body: A Cultural Analysis of Reproduction*. Boston: Beacon Press.
Martin, Paul, Richard Ashcroft, George T. H. Ellison, Andrew Smart, and Richard Tutton. 2007. *Reviving "Racial Medicine"? The Use of Race/Ethnicity in Genetics and Biomedical Research, and the Implications for Science and Healthcare*. London: Faculty of Health and Social Care Sciences at St. George's, University of London.
Messer, A'ndrea Elyse. 2014. "Scientific Racism's Long History Mandates Caution." *Penn State News*, February 14. Retrieved August 19, 2014 (http://news.psu.edu/story/304151/2014/02/14/research/scientific-racisms-long-history-mandates-caution).
Morning, Ann. 2001. "The Racial Self-identification of South Asians in the United States." *Journal of Ethnic and Migration Studies* 27(1):61-79.
Morning, Ann. 2008. "Ethnic Classification in Global Perspective: A Cross-national Survey of the 2000 Census Round." *Population Research and Policy Review* 27(2):239-72.
Morning, Ann. 2011. *The Nature of Race: How Scientists Think and Teach about Human Difference*. Berkeley: University of California Press.
Nelkin, Dorothy and M. Susan Lindee. 1995. *The DNA Mystique: The Gene as Cultural Icon*. New York: Freeman.
Nelson, Alondra. 2008. "Bio Science: Genetic Genealogy Testing and the Pursuit of African Ancestry." *Social Studies of Science* 38(5):759-83.
Nobles, Melissa. 2000. *Shades of Citizenship: Race and the Census in Modern Politics*. Stanford, CA: Stanford University Press.
Obasogie, Osagie K. 2014. *Blinded by Sight: Seeing Race through the Eyes of the Blind*. Stanford, CA: Stanford University Press.
Omi, Michael and Howard Winant. 1986. *Racial Formation in the United States*. New York: Routledge and Kegan Paul.
Omi, Michael and Howard Winant. 1994. *Racial Formation in the United States: From the 1960s to the 1990s*. New York: Routledge.
Ossorio, Pilar and Troy Duster. 2005. "Race and Genetics: Controversies in Biomedical, Behavioral, and Forensic Sciences." *American Psychologist* 60(1):115-28.
Paschou, Peristera, Elad Ziv, Esteban G. Burchard, Shweta Choudhry, William Rodriguez-Cintron, Michael W. Mahoney, and Petros Drineas. 2007. "PCA-correlated SNPs for Structure Identification in Worldwide Human Populations." *PLoS Genetics* 3(9):1672-86.
Pritchard, Jonathan K., Matthew Stephens, and Peter Donnelly. 2000. "Inference of Population Structure Using Multilocus Genotype Data." *Genetics* 155(2):945-59.
Risch, Neil, Esteban Burchard, Elad Ziv, and Hua Tang. 2002. "Categorization of Humans in Biomedical Research: Genes, Race and Disease." *Genome Biology* 3(7):1-12.
Roberts, Dorothy. 2011. *Fatal Invention: How Science, Politics, and Big Business Re-create Race in the Twenty-first Century*. New York: New Press.
Rosenberg, Noah A., Saurabh Mahajan, Sohini Ramachandran, Chengfeng Zhao, Jonathan K. Pritchard, and Marcus W. Feldman. 2005. "Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure." *PLoS Genetics* 1(6):660-71.
Rosenberg, Noah A., Jonathan K. Pritchard, James L. Weber, Howard M. Cann, Kenneth K. Kidd, Lev A. Zhivotovsky, and Marcus W. Feldman. 2002. "Genetic Structure of Human Populations." *Science* 298:2381-85.
Rubin, Gayle. 1975. "The Traffic in Women: Notes on the 'Political Economy' of Sex." Pp. 157-210 in *Toward an Anthropology of Women*, edited by R. Reiter. New York: Monthly Review Press.
Samhan, Helen Hatab. 1999. "Not Quite White: Race Classification and the Arab-American Experience." Pp. 209-26 in *Arabs in America: Building a New Future*, edited by M. W. Suleiman. Philadelphia, PA: Temple University Press.
Sanders, Edith R. 1969. "The Hamitic Hypothesis: Its Origins and Functions in Time Perspective." *Journal of African History* 10(4):521-32.
Sarich, Vincent and Frank Miele. 2004. *Race: The Reality of Human Differences*. Boulder, CO: Westview Press.
Satel, Sally. 2002. "I Am a Racially Profiling Doctor." *The New York Times Magazine*, May 5, pp. 56-58.

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014

Schiebinger, Londa. 1993. *Nature's Body: Gender in the Making of Modern Science*. Boston, MA: Beacon Press.

Serre, David and Svante Pääbo. 2004. "Evidence for Gradients of Human Genetic Diversity within and Among Continents." *Genome Research* 14:1679-85.

Shah, Anish M., Rakesh Tamang, Priya Moorjani, Deepa Selvi Rani, Periyasamy Govindaraj, Gururaj Kulkarni, Tanmoy Bhattacharya, Mohammed S. Mustak, L.V.K.S. Bhaskar, Alla G. Reddy, Dharmendra Gadhvi, Pramod B. Gai, Gyaneshwer Chaubey, Nick Patterson, David Reich, Chris Tyler-Smith, Lalji Singh, and Kumarasamy Thangaraj. 2011. "Indian Siddis: African Descendants with Indian Admixture." *American Journal of Human Genetics* 89(1):154-61.

Shapin, Steven. 1995. "Here and Everywhere: Sociology of Scientific Knowledge." *Annual Review of Sociology* 21:289-321.

Shiao, Jiannbin Lee, Thomas Bode, Amber Beyer, and Daniel Selvig. 2012. "The Genomic Challenge to the Social Construction of Race." *Sociological Theory* 30 (2):67-88.

Smedley, Audrey and Brian Smedley. 2012. *Race in North America: Origin and Evolution of a Worldview*. Boulder, CO: Westview Press.

Stepan, Nancy. 1982. *The Idea of Race in Science: Great Britain 1800-1960*. London: Archon Books.

Templeton, Alan R. 1999. "Human Races: A Genetic and Evolutionary Perspective." *American Anthropologist* 100(3):632-50.

U.S. Office of Management and Budget. 1997. *Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity*. Washington, DC: Author.

Wailoo, Keith. 2001. *Dying in the City of the Blues: Sickle Cell Anemia and the Politics of Race and Health*. Chapel Hill: University of North Carolina Press.

Wailoo, Keith and Stephen Pemberton. 2006. *The Troubled Dream of Genetic Medicine: Ethnicity and Innovation in Tay-Sachs, Cystic Fibrosis, and Sickle Cell Disease*. Baltimore: Johns Hopkins University Press.

Weber, Max. [1956] 1978. *Economy and Society: An Outline of Interpretive Sociology*. Berkeley: University of California Press.

Wilson, James F., Michael E. Weale, Alice C. Smith, Fiona Gratrix, Benjamin Fletcher, Mark G. Thomas, Neil Bradman, and David B. Goldstein. 2001. "Population Genetic Structure of Variable Drug Response." *Nature Genetics* 29(3):265-69.

Winant, Howard. 2001. *The World Is a Ghetto: Race and Democracy Since World War II*. New York: Basic Books.

AUTHOR BIOGRAPHY

Ann Morning is an associate professor of sociology at New York University and a faculty affiliate of New York University Abu Dhabi. Her research interests include race, demography, and the sociology of science, especially as they pertain to census classification worldwide and to scientific and lay concepts of racial difference. She is the author of *The Nature of Race: How Scientists Think and Teach about Human Difference* (2011).

Downloaded from stx.sagepub.com at UNIV OF WISCONSIN on October 15, 2014