

Are Clusters Races? A Discussion of the Rhetorical Appropriation of Rosenberg et al.'s "Genetic Structure of Human Populations"

Melissa Wills*

Noah Rosenberg et al.'s 2002 article "Genetic Structure of Human Populations" reported that multivariate genomic analysis of a large cell line panel yielded reproducible groupings (clusters) suggestive of individuals' geographical origins. The paper has been repeatedly cited as evidence that traditional notions of race have a biological basis, a claim its authors do not make. Critics of this misinterpretation have often suggested that it follows from interpreters' personal biases skewing the reception of an objective piece of scientific writing. I contend, however, that the article itself to some degree facilitates this misrepresentation. I analyze in detail several verbal and visual features of the original article that may predispose aspects of its racial interpretation; and, tracing the arguments of one philosopher and one popular science writer, I show how these features are absorbed, transformed into arguments for a biological basis of race, and re-attributed to the original. The essay demonstrates how even slight ambiguities can enable the misappropriation of scientific writing, unintentionally undermining the authors' stated circumspection on the relationship between cluster and race.

Keywords

race • genetics • *structure* • clusters • populations • popular science

1 Introduction

In their paper "Genetic Structure of Human Populations," published in *Science* in 2002, Noah Rosenberg et al. reported that the software program *structure* was able to assign individuals from diverse human populations into reproducible groupings based on their genotypes at hundreds of loci and, further, that these clusters generally aligned with individuals' regional and continental origins. The article has been tremendously influential: it was the most-cited paper in *Science* in its year and continues to feature in scientific, philosophical, and cultural debates fifteen years

*Department of English, Voorhies Hall, One Shields Avenue, University of California, Davis, CA 95616, USA, mawills@ucdavis.edu

Received 17 February 2016; Revised 20 June 2017; Accepted 23 June 2017
doi:10.3998/ptb.6959004.0009.012



later. Much of its treatment in the media has been controversial, with some interpreters describing the study as proof of a biological grounding for conventional notions of race, even though the article's report of clusters approximating the five continents traditionally identified as the origins of the major races is not the only—or the primary—result of the study. The authors, for their part, have declined to suggest any relationship between world-level clustering data and social categories of race. The article itself never once uses the word “race,” with patterns referred to only as “clusters,” “groups,” or “populations.” And it sidesteps discussions of race altogether, considering the results only in terms of possible implications for epidemiology and evolutionary history. Elsewhere, Rosenberg et al. (2005) have more explicitly denied any potential racial implications of their research, specifying that it “should not be taken as evidence of our support of any particular concept of ‘biological race’” (0668–9).

This essay, while assuming the authors to be neutral on matters of race, asks how a race-oriented interpretation might arise from the article's text and figures. I argue that the study is presented with subtle linguistic and visual ambiguities that potentially predispose a reader toward the very interpretation that the authors deny. I begin my discussion in the following section with a more detailed overview of the article's figures and conclusions, before turning in section 3 to an assessment of its continuing currency in philosophical and popular science discourses on race. The remaining sections analyze various facets of the paper's visual and textual rhetoric. Section 4 draws attention to its foregrounding of the world-level, continental clusters featuring prominently in race debates. Section 5 shows how the text and figures promote an impression of coherent, distinct clusters, and section 6 details the article's reliance on a specific word choice that suggests a particularly tight relationship between cluster and geography. Sections 4–6 work closely with two particular misappropriations of the study, showing how they similarly seize and transform each facet into supposed proof of the biological basis of race. Section 7 concludes the essay with a reflection on scientific discourse and its cultural interpretations.

2 Study Overview

Before assessing the paper's rhetoric, a brief overview of its methods and findings is warranted. Rosenberg et al. (2002) conducted their study using the clustering algorithm *structure*, which postulates population structures based on genotype data from multiple individuals without prior knowledge of their populations of origin.¹ K , the number of clusters, is variable and user-specified such that once K is set, allele frequencies across many loci are used to identify that number of clusters and to define cluster memberships for individuals. An innovation of *structure* was its ability to assign individuals partial cluster membership (given as membership coefficients summing to 1.0 across all clusters), enabling the recognition of admixture and a more fine-grained picture of population structure.

Rosenberg et al. applied *structure* to the HGDP-CEPH Human Genome Diversity Cell Line Panel, drawn from 1,056 individuals from 52 distinct global populations.² The analysis was based on genotype data at 377 microsatellite loci, or noncoding regions of DNA characterized by short, repeating nucleotide motifs. The researchers found that membership coefficient patterns tended to be consistent across individuals within predefined populations and further, that inferred clusters agreed to some extent with geographical regions at both the world and regional levels. The authors suggest that population structures are likely attributable to genetic drift, with geographical or linguistic isolation driving genetic differentiation between groups.

¹*structure* was developed by Pritchard et al. (2000) at Oxford.

²For details of the cell line panel, see Cann et al. (2002).

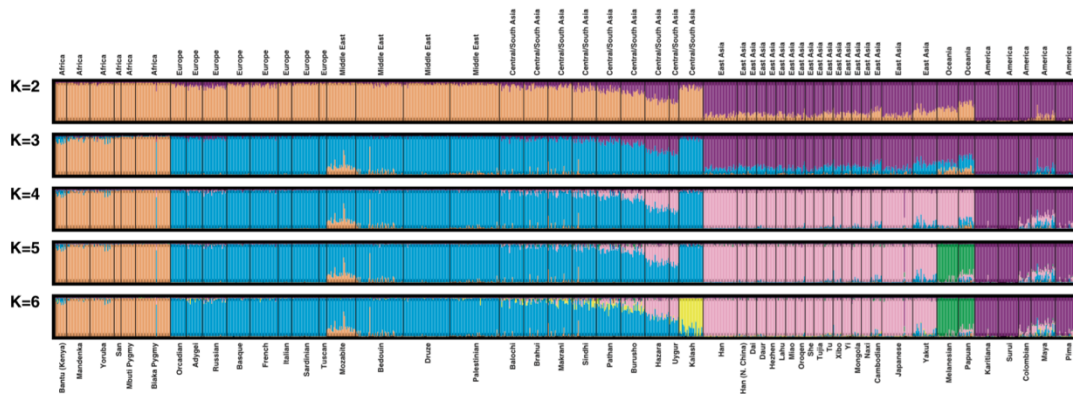


Fig. 1. Estimated population structure. Each individual is represented by a thin vertical line, which is partitioned into K colored segments that represent the individual's estimated membership fractions in K clusters. Black lines separate individuals of different populations. Populations are labeled below the figure, with their regional affiliations above it. Ten *structure* runs at each

K produced nearly identical individual membership coefficients, having pairwise similarity coefficients above 0.97, with the exceptions of comparisons involving four runs at $K = 3$ that separated East Asia instead of Eurasia, and one run at $K = 6$ that separated Karitiana instead of Kalash. The figure shown for a given K is based on the highest probability run at that K .

Figure 1: From 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: 2381–2385. Reprinted with permission from AAAS.

The evidence for these conclusions is presented chiefly in two brightly colored figures. In each, *structure's* results are visualized by assigning each cluster a different color, with each individual represented by a thin vertical line divided into colored segments proportional to inferred membership coefficients for the number of clusters being assessed.³ Individuals are grouped with others of their population of origin and ordered according to their Cann et al. (2002) identification numbers. Clusters, then, emerge as swatches of color spanning horizontal bands of aggregated individual lines, concentrating in certain areas and diffusing in others.

Figure 1, reproduced here, visualizes world-level clustering patterns across all 1,056 individuals for K values ranging from 2 through 6. Populations are placed according to continent of origin, labeled with population names across the bottom and continental affiliations across the top. The article points to Figure 1 to support its conclusions that clustering patterns are generally consistent within populations and that broader patterns of cluster membership generally “correspond” to geographical divisions. At $K=6$, for instance, Africa appears predominantly orange, America purple, and so on.

Figure 2, also reproduced here, displays further clustering results for the groups comprising each major continent, with Eurasia additionally segmented into its component regions. K values are more variable here, with the researchers including data for the highest values of K yielding consistent results in each region. The article's discussion describes in these results a general consistency of membership coefficients within populations as well as a tendency for clusters to coincide with geographical regions.

Rosenberg et al. conclude the article by speculating on some possible mechanisms for the patterns detected, specifically genetic drift resulting from historical migrations and linguistic isolation. They also comment on potential applications of their findings in epidemiology and medicine. The article ends with a reminder of the overall genetic similarity of human populations, noting that “because most alleles are widespread, genetic differences among human populations derive mainly from gradations in allele frequencies rather than from distinctive ‘diagnostic’ genotypes. Indeed, it was only in the accumulation of small allele-frequency differ-

³The code used to generate Figures 1 and 2 was eventually adapted into the software program *Distrupt*; see Rosenberg (2007a, 2007b).

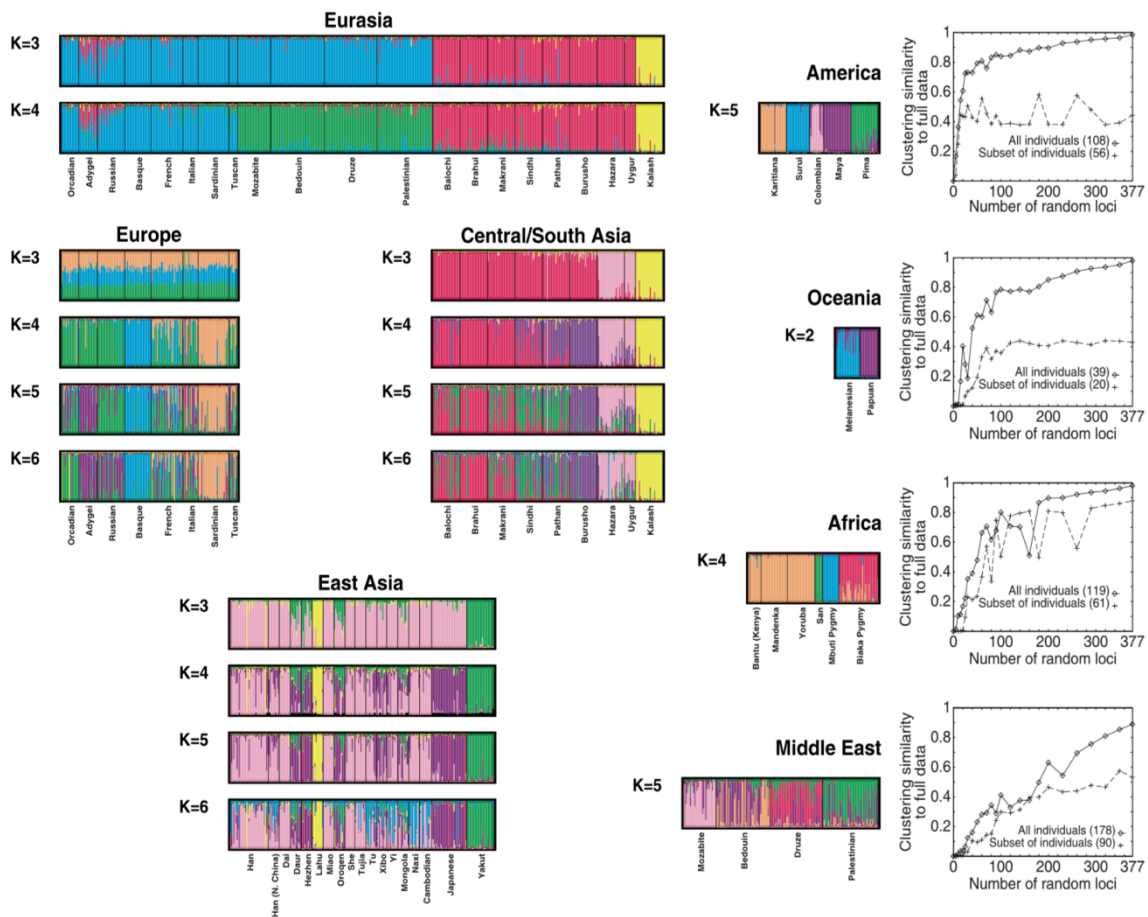


Fig. 2. Estimated population structure for regions. For America, Oceania, Africa, and the Middle East, solutions were consistent across 10 runs (all similarity coefficients above 0.97, 0.93, 0.97, and 0.86, respectively, except those involving one run with Africa that assigned many Biaka individuals partial membership with San). Values of *K* shown for these samples are the highest values for which this was true, and the highest

probability runs are shown. For remaining regions, solutions were more variable across runs, and the highest probability runs for various values of *K* are displayed. Graphs for America, Oceania, Africa, and the Middle East display median similarity coefficients between runs based on the full data and runs based on subsets of the data. Correspondence of colors across figures for different regions is not meaningful.

Figure 2: From 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: 2381–2385. Reprinted with permission from AAAS.

ences across many loci that population structure was identified” (2384).

3 Context & Methods

3.1 Race Ontologies

As has been previously rehearsed in philosophy of science, differences in beliefs about the reality of race can be sorted according to the types of ontological claims that they entail: realist, anti-realist, or constructivist.⁴ The present article is not an effort to develop an ontological account of race as a biological or cultural concept; rather, it interrogates two distinct types of racial realist claims. These claims can be distinguished according to the categories of racial realism delineated by Kaplan and Winther (2014), two of which relate to biological categories: *bio-genomic cluster/racial realism*, which claims that population structure exists in humans, as assessed through genomic or anthropometric measures; and *biological racial realism*, which claims that groups identified genomically or phenomically map stably onto social groups conventionally identified as races (1040–1).⁵ As this section will demonstrate, Rosenberg et al. (2002) embodies the first of these stances, while two of its most contentious interpretations align far more radically with the second stance. I contend in what follows that this shift, in which the study is exaggerated and distorted into proof of a genetic basis for social categories of race, is mediated in part by its own distinct verbal and visual features.

Rosenberg et al.’s article articulates a bio-genomic cluster/racial realist stance in expressing a claim that human population structures exist among the individuals sampled and that these are discernible through *structure*’s genomic analysis, while declining to suggest any relationship between such structures and socially-defined races. Whereas other population genetics studies (e.g., Tang et al. 2005) specifically interrogate the relationship between genomically-identified population structure and self-identified racial identity, Rosenberg et al.’s study neither tests nor posits any such correspondence, noting only a “general agreement of genetic and predefined [geographical] populations” (2381). Their subsequent disavowals of racial significance, as noted above, affirm Kaplan’s and Winther’s observation that “this kind of realism is not necessarily about a ‘race’ concept ... some influential and socially responsible population geneticists have no desire to become involved in debates over race” (1040). To emphasize this discontinuity and more accurately reflect the scope of the research study, in the remainder of this article I simplify Kaplan’s and Winther’s term to *bio-genomic cluster realism*.⁶

The study’s refusal to engage with questions of race, however, does not prevent it from being taken to be about race, as is done by the authors whose claims I detail below. I have selected two writers whose work—in different fields, for different audiences, four years apart—has sparked criticism for falsely representing Rosenberg et al. (2002) in attempting to substantiate an altogether different realist ontology. Both interpreters attribute an outsized influence to the study, representing it as foundationally important to the recognition that races are biologically grounded. Specifically, each writer asserts a biological racial realist stance, falsely claiming that the study reveals an equivalence between clusters and the social groups identified as races (Kaplan and Winther 2014, 1040). In focusing on these two interpretations, I show how Rosenberg

⁴Kaplan and Winther (2013, 2014) have revived these categories, drawing on Mills (1988).

⁵The third type, *social racial realism*, which defends race as a legitimate social category regardless of biology, will not be discussed here (2014, 1041).

⁶Kaplan and Winther note that their term *bio-genomic cluster/race realism* includes a forward slash to indicate that, while this kind of realism can put forth a race concept, it does not necessarily do so (2014, n. 1).

et al.'s paper has survived in a legacy that exceeds the disciplinary boundaries and realist ontology within which it was written.

First is Neven Sesardic, whose (2010) article marshals a number of population genetics studies in an attempt to challenge what he sees as a dominant scientific consensus that race is a social construction divorced from any biological reality. He describes Rosenberg et al. (2002) as a pivotal refutation of that consensus, heralding a decisive paradigm shift as “an important discovery that makes it much more difficult than before to claim that race is entirely disconnected from genetics” (153). Sesardic’s hedged language here belies his audacious claim that the Rosenberg et al. study singlehandedly demolishes constructivist arguments. He later writes derisively that such authors should “make contact with the most recent exciting developments in genetics and deal with the best contemporary attempts to rehabilitate the biological foundations of race” (154). A later (2013) formulation is even more explicit, citing subsequent clustering analyses that Sesardic claims amount to a “scientific validation of the race concept” (291).

Sesardic’s views have not gone unchallenged. Taylor (2011) has argued that his claims about distinguishing genetic groups are insufficient to rehabilitate a biological race concept; Hochman (2013, 2016) has dismantled the logical inconsistencies in what he terms “one of the strongest defences of racial naturalism in recent times” (2013, 278); and Pigliucci (2013) critiques Sesardic on both scientific and philosophical grounds. I build from these analyses in my discussion below.

The study’s findings are recruited similarly in a more recent text, Wade’s (2014) popular press book. Like Sesardic, Wade targets a perceived orthodoxy defining race as a merely social phenomenon, yet his biological racial realist stance goes beyond Sesardic’s in arguing pervasively that the ongoing process of natural selection has conferred “not only expected traits like skin color and nutritional metabolism but also some aspects of brain function” (4). The claim is distinctly hereditarian and is provocatively justified in the book’s fifth chapter, “The Genetics of Race,” which grants the Rosenberg et al. study prime significance as a foundational moment in population genetics.⁷ Wade’s interpretation draws no distinction between cluster and race, assuming a perfect mapping of the two from which the book’s later claims about racial heritage and physiology devolve. His over-interpretation of this and similar studies claims that they reveal racial implications that are absent from the originals.

Wade’s work has been extensively criticized. In addition to numerous individual responses, a group of 143 scientists—including Rosenberg and several of his co-authors—signed a letter to the editor of *The New York Times* condemning the book’s assertions and concluding that “there is no support from the field of population genetics for Wade’s conjectures” (Coop et al. 2014).⁸ Criticism has abounded among nonspecialist writers as well. For instance, Orr’s extensive and prominent (2014) review argues that the book “goes beyond reporting scientific facts or accepted theories and finds Wade championing bold ideas that fall outside any scientific consensus.” Despite my critiques of certain aspects of Rosenberg et al.’s article below, I fully concur with these criticisms, though my analysis is focused more narrowly on Wade’s abuse of this one particular study.

Some critics have expressed concern that such misinterpretations of Rosenberg et al. (2002) are prevalent. Bolnick suggests that the article has been “widely cited as verifying traditional ideas about race and the pattern of human biological diversity” (2008, 74). Similarly, Templeton writes that many authors referencing the article have “claimed that this paper supported the idea that races were biologically meaningful in humans” (2013, 267). While neither Bolnick

⁷Hereditarian positions are a subset of biological racial realisms, in Kaplan’s and Winther’s classification (2014, 1041).

⁸Individual scientific responses have included Jobling (2014) and Feldman (2014), among many others.

nor Templeton fully substantiates this trend, and it is beyond the scope of this essay to conduct a survey of such misrepresentations, their warnings underscore the importance of asking how such a pattern might have arisen.⁹ My task below is to detail the work of misapprehension in Sesardic and Wade, showing how certain features of the original article have been transformed into evidence for claims that its authors have never made.

The racial interpretation as it appears in Sesardic and Wade is driven by three central moves, relating to the three key features of the original article discussed below. I define it as follows: The racial interpretation of “Genetic Structure of Human Populations” (i) takes as the object and primary result of the study the detection of five large clusters in the worldwide sample; (ii) overemphasizes the uniformity and distinctiveness of clusters; and (iii) asserts a direct mapping between these clusters and the five traditionally identified continental races, implying that clusters can serve as proxies for generalizations about skin color or other racial phenotypes.

3.2 *Methodology and Critical Context*

Defenders of Rosenberg et al. (2002) have often sought to correct the racial misinterpretation by reemphasizing the aims and findings of the original article. Rebuttals have tended, for instance, to reject the persistent equation of cluster and race by reiterating the study’s refusal to draw this connection. Likewise, they have generally disputed the overemphasis on the detection of five world-level clusters by clarifying that the study actually found a variable number of clusters and is therefore less conclusive—“more cautious”—than it is made to appear (Templeton 2013, 267).¹⁰ Such errors are often attributed to the interpreters’ preexisting racial biases. Bolnick, for example, writes that the five worldwide clusters have been so emphasized “simply because they fit the general notion in our society that continental groupings are biologically significant. This notion is a legacy of traditional racial thought and seems to persist even when not clearly supported by biological data” (2008, 77).

While such rebuttals are vital in rejecting the racial interpretation, I challenge this tendency to scrutinize the language and motives of the interpreters without considering their source material. The influence of personal biases in the reporting of research relevant to the race debate surely cannot be ignored; like the article’s defenders, I accept the researchers’ disavowal of racial significance and place the primary responsibility for the racial interpretation on the interpreters themselves.¹¹ However, to emphasize only errors of interpretation risks ceding too much authority to the scientific work, blocking discussion of how these misappropriations might have been prevented in the first place. My primary aim is therefore to redirect attention to the article itself, focusing on a number of rhetorical features that are directly imported into Sesardic’s and Wade’s arguments and which to some degree enable its misrepresentation. These features are verbal and visual, distinguishing my approach from those of authors who have critiqued the study’s design or the nature of clustering methods more generally.¹²

⁹In sociology, Shiao et al. (2012) put the article to similar ends as Wade and Sesardic, in an argument that draws on Sesardic (2010) and has been extensively challenged by Fujimura et al. (2014).

¹⁰Similar rebuttals include Bolnick (2008), Fujimura et al. (2014), and Pigliucci (2013).

¹¹I distinguish this approach from critiques citing Rosenberg et al. (2002) as evidence of racially-biased study design and/or more systemic bias in molecular genetics, e.g., Yudell (2014) and Duster (2003). Because these critiques cite the article dismissively and without any detailed analysis—and because they sometimes (as in Yudell) falsely represent the study as testing correspondence between cluster and race—I sidestep them entirely.

¹²Scientific critiques of Rosenberg et al.’s methods, or clustering methods generally, have included Serre and Pääbo (2004) and Templeton (2013), among others. Humanities and social science critiques

My work in the remaining sections is grounded in the discipline of science rhetoric and thus in the close analysis of scientific writing's textual features. Building from classical predecessors such as Bazerman's (1988) study of the experimental research article and Myers' (1990) analysis of scientific discourse across genres, I approach Rosenberg et al. (2002) with "the kind of detailed attention usually reserved for literature" (Myers 1990, ix). I scrutinize significant word choices and phrasings, finding openings for the racial interpretation even in the sparse prose of scientists who wish not to discuss race. My readings of Sesardic and Wade also examine the one-step discursive shift—"accommodation" (Fahnestock 1993)—in which a scientific message is reinterpreted for non-specialist audiences. My close-reading analysis demonstrates how the article's racial interpreters recruit and transform features of the original article in support of their ideologies.

However, classical science rhetoric alone—which adheres almost exclusively to textual features of argumentation—is insufficient for analyzing the article at hand. The persuasive power of Rosenberg et al. (2002) resides chiefly in its two figures, and so these too must be assessed with rhetorical strategy in mind. Abundant interdisciplinary scholarship is devoted to scientific visuals, their epistemic functions, design theory, and the cognitive processes of their apprehension. My work is informed by these approaches, but my primary objective is to show how Rosenberg et al.'s images function argumentatively. There is extensive scholarly precedent for considering scientific visuals within their social contexts, showing how they are adapted to suit their creators' varying theoretical commitments (e.g., Gannett and Griesemer 2004b), disciplinary conventions (e.g., Taylor and Blum 1991), and pedagogical aims (e.g., A. Gross 2012). Gannett and Griesemer, for example, write of mapping practices that

Maps are abstractions that privilege some features of the world and ignore others. Maps are perspectival—what gets included on a map and what gets left off of a map depends on the aims, needs, interests, and conventions associated with specific incidences of map-making. (2004b, 75)

I follow these approaches in my detailed visual analysis, taking the article's figures to be constructed to meet certain rhetorical requirements and asking how their specific construction—the placement of labels, the ordering of boxes—creates particular impressions.

The analysis below is equally focused on the broader argumentative context to which these figures belong, that is, their function within the article as a whole. Gross and Harmon (2014) merge verbal and visual rhetorical analysis in a comprehensive account of scientific argumentation, asserting that "understanding the images appearing within scientific texts involves not only interpreting the meaning behind the pattern within a given visual, but also assimilating that meaning into an argument or narrative" (46).¹³ I follow Gross and Harmon's lead below, considering the article's figures as continuous with its textual argumentation in a unified persuasive task. I therefore consider its visual features alongside its verbal ones, moving from topic to topic (the major subsections) and conducting both types of analysis within each to forge a cohesive whole.

I wish to clarify before proceeding that my analysis, while critical, is not accusatory. The rhetorical elements I discuss in Rosenberg et al. (2002) are assumed to be incidental, reflecting the discourse of the race-charged society within which the article is embedded rather than indicating any particular racial bias amongst the scientists. In contrast, I consider the hardline racial interpretations of Sesardic and Wade to originate chiefly with these authors themselves.

have included Bolnick (2008) and Weiss and Fullerton (2005).

¹³Cambrosio et al. (2005) provides another account of verbal-visual interaction.

My argument does not seek to blame the researchers but rather to demonstrate the difficulty of presenting population genetics research in a way that fully defends against these interpretations, particularly in the extremely space-limited format of a *Science* Report. In the work of these two racially-motivated interpreters, even the smallest rhetorical features have been extracted, misrepresented, and re-attributed to the original. It is through the fine details of language and image that the study becomes recast as something it never was.

4 Worldwide $K=5$ Data as Primary

As noted above, the racial interpretation of Rosenberg et al. (2002) begins from point (i): taking as the study's objective and primary finding the detection of five large clusters in the worldwide sample, when in fact the study identifies a variable number. In this section, I discuss how the article sets the precedent for this conclusion through a number of features subtly privileging the world-level population data presented in Fig. 1—and the $K=5$ data in particular—over its other findings. I focus my analysis on two primary sites, the abstract and the presentation of worldwide clustering data, as well as a particular omission, and conclude by showing how Sesardic and Wade propagate this emphasis on five worldwide clusters.

4.1 Abstract

As early as the abstract, the authors seem to identify five continental clusters as the primary finding of the study, observing that “we identified six main genetic clusters, five of which correspond to major geographic regions, and subclusters that often correspond to individual populations” (2381). Bolnick has noted the discrepancy between this statement's mention of six clusters and the more expansive analysis actually described within the paper (2008, 76). I would add that in overtly identifying the six worldwide clusters of Figure 1 as the “main” clusters detected, the statement further suggests that the regional population structures depicted in Figure 2 are subsidiary. This impression is reinforced by the differing degrees of certainty assigned to each: the continental clusters are said to “correspond to major geographic regions,” while the regional clusters only “often correspond” to individual populations. Admittedly, Figure 2's lower-resolution clustering patterns to some degree force this language, but the greater certainty verbally accorded to the world-level data in the abstract seems to grant Figure 1 precedence. Moreover, the primacy of the worldwide data is reinforced in the use of the word “subclusters” to refer to the patterns distinguished in the regional analysis. This word choice suggests that the clusters in Figure 2 are somehow derived from those identified in Figure 1—that they are refinements of the $K=5/6$ clusters—when in reality they reflect separate *structure* analysis performed on subsets of individuals. Adjusting this terminology, perhaps replacing “subclusters” with “sub-regional clusters” or “intra-regional clusters,” might have helped to prevent this ambiguity.¹⁴

With the worldwide data, the authors' phrasing further implies a particular priority for the $K=5$ over the $K=6$ data. The phrase quoted above mentions the detection of six clusters before immediately distinguishing the five that correspond to major geographic regions, in what could be seen as a dismissal of the sixth group (the Kalash population), refocusing the reader on the unspecified alignment of clusters with continental origins. Any such implication is transient, as the Kalash population is indeed addressed within the body of the paper. However, the point bears mentioning here given that Wade and Sesardic fixate so narrowly on the $K=5$, rather than $K=6$ data. Since abstracts are taken, particularly by nonspecialist readers, to summarize major

¹⁴I thank an anonymous reviewer for formulating the first of these alternatives.

that any of the reported clustering patterns best reflects the actual number of human populations present in the sample. At the same time, however, Rosenberg et al. offer no guidance as to how to interpret the various values of K , an omission that in combination with the factors discussed above may predispose the reader to overlook other aspects of the study.

There are legitimate reasons for avoiding claims as to the “best” value of K . As Rosenberg et al. note in their supplementary materials, one approach to inferring K would be to run *structure* across numerous values of K , then select the one maximizing the posterior probability of the data (S1). They observe, however, that in complex data sets *structure* analysis can produce multiple clustering schemes and therefore different estimated probabilities for a given K , making comparison difficult.¹⁶ In addition, the utility of *structure*’s statistical calculations has itself been contentious from the outset. Pritchard et al. (2000), in introducing the algorithm’s methodology, observed that “the problem of inferring the number of clusters, K , present in a data set is notoriously difficult,” involving “severe computational difficulties” and producing only approximate values (949).¹⁷ Given these difficulties, Rosenberg et al. include results for multiple values of K in parallel for each figure, with each horizontal bar reflecting the highest probability run for that K . This presentation strategy prioritizes no particular K and conveys no inferred population number. Bolnick suggests that this approach justifies a wider view of the article’s scope than is present in its more reductive interpretations, observing that the $K=6$ data “does not necessarily provide a better representation of human genetic differentiation than the clustering observed when K is set to 4, 9, 12, or any other number” (2008, 76).

Yet even though the authors conscientiously decline to present claims or data concerning the number of populations represented, these omissions are not addressed in the discussion. The reader is given little direction as to how to compare results across multiple values of K , or about the relationship of K to actual human populations. While such guidance would likely not be needed for an audience of population geneticists familiar with the underlying assumptions and methods, for lay readers—particularly those motivated to look for evidence of a biological basis of race—this oversight might well have unwanted side effects. A brief discussion of K ’s variable status in the context of the study’s scope and limitations might have hindered, at least somewhat, the racial interpretation from emerging.

4.4 Rhetorical Intensification

The implication that Rosenberg et al. (2002)’s major finding was the appearance of five worldwide clusters is foundational to the racial interpretation’s shift from bio-genomic cluster realism to biological racial realism. Sesardic’s description of the study singles out the $K=5$ data, stating that *structure* “did allow an inference of group structure and that, furthermore, five clusters derived from that analysis of purely genetic similarities correspond largely to major geographic regions This is an important discovery” (2010, 153). The second half of his statement reproduces the abstract almost verbatim, with a further reduction: whereas the abstract verbally minimizes the sixth worldwide cluster, Sesardic neglects to mention it at all. He presents the $K=5$ data as the sole finding of the study, and relies on unstated conventional associations between race and geography to conclude that the study establishes a link between cluster and

¹⁶Evanno et al. (2005) have observed that *structure*’s method of using $Pr(X|K)$ to differentiate between K ’s does not work well in practice. They propose an alternative statistical method for inferring population number that seems to work better, but this method was not yet available to Rosenberg et al. in 2002.

¹⁷Pritchard et al. (2000) explain these difficulties in depth, as does Bolnick (2008) within a larger context of ancestry inference studies including Rosenberg et al. (2002).

race.¹⁸ Pigliucci has described this synopsis as “an exercise in selective quotation” because of its exclusive focus on the $K=5$ data, asking “why pick a particular [number of clusters] as the major finding of the paper, other than because five clusters happen to fit the author’s predilection for the true number of human races?” (2013, 273).

Wade similarly identifies the worldwide $K=5$ as the study’s primary finding, though his description is more charged. He writes with a tone of assured self-evidence that the study “showed, as expected, that the 1,000 individuals ... clustered naturally into five groups, corresponding to the five continental races” (2014, 97–8). Like Sesardic, he appears to adopt the formulation presented in the paper’s abstract in focusing on $K=5$ as the finding of interest, to the exclusion of any other results. But whereas Sesardic remains more hedged throughout his essay about the relationship between cluster and race, Wade is overt in asserting that those five groups *are*—“naturally,” “as expected”—the same as the continental races.

Claims like these do not automatically follow from the paper itself, even though as I have argued in this section it skews toward a privileging of this particular finding. Nevertheless, these interpretations show how readily even the slightest prioritization of five clusters can be taken to signify five biological races. One can see in retrospect just how little is needed to set off this chain of interpretations, and thereby the need for abundant disclaimers as to the concreteness of the results.

5 Cluster Uniformity and Distinctiveness

The second key component of the racial interpretation of “Genetic Structure of Human Populations” is (ii): overemphasizing the uniformity and distinctiveness of clusters. This section argues that the original article may subtly predispose this maneuver through the arrangement and discussion of its two figures. I focus on the appearance of uniform clusters within population groups in both figures and of sharp divisions between worldwide clusters in Figure 1. I conclude by showing how Sesardic and Wade expand on these patterns.¹⁹

5.1 Population and Regional Uniformity

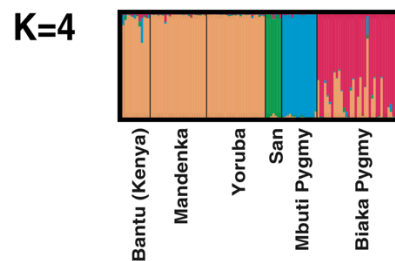
Rosenberg et al. repeatedly emphasize similarity in cluster membership within populations, regions, and continents; for instance, they note that “in the worldwide sample, individuals from the same predefined population nearly always shared similar membership coefficients in inferred clusters” and that within populations they found “similar membership coefficients for most individuals” (2382). General similarity is indeed a major conclusion of the study, affirming that population structure is detectable by the algorithm. I argue, however, that the paper occasionally overemphasizes the uniformity of clusters, subtly de-emphasizing heterogeneity and unwittingly encouraging a geographical determinism that feeds into the racial interpretation.

Despite the overall pattern of population and regional coherence, the article’s figures occasionally reveal more diversity at close range than is first apparent. For example, the African Biaka Pygmy population in Figure 2 is largely colored red in what seems a distinction from neighboring populations. Yet the group appears much more diverse when magnified, with the

¹⁸Sesardic (2013) has widened his interpretation in response to Hochman (2013), no longer prioritizing any particular number of clusters; Hochman’s response (2016) notes this move is incompatible with Sesardic’s account of racial naturalism (66).

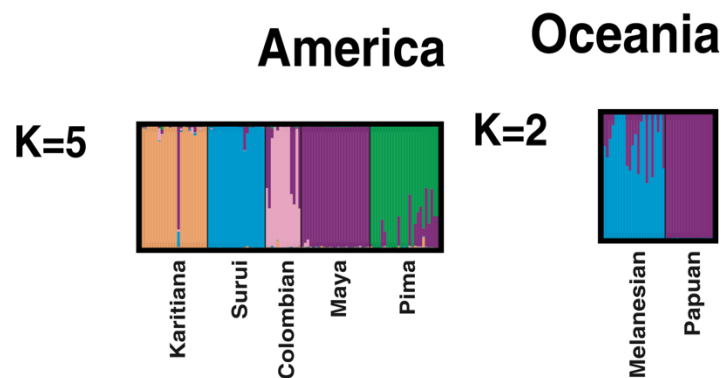
¹⁹My approach is similar to Gannett’s and Griesemer’s analysis of human blood type frequency bar graphs, in which the exaggeration of between-group differences and the minimization of within-group differences are taken to imply a judgment of some kind (2004a, 134–5).

prominent red coloration disrupted by a number of individuals bearing extensive membership in the orange cluster, and lesser but visible membership in the blue and green clusters:



These secondary structures are easy to miss without zooming in, despite representing roughly 20% of the total area, and would likely be more visible if the lines were grouped according to the pattern of cluster membership. While the placement of individuals within groups was not determined by Rosenberg et al., the scattering of secondary patterns nevertheless has the effect of making the group appear more coherent—more essentially red—than it really is.²⁰ That impression is not dispelled by the article’s text, in which the authors acknowledge that the inferred African clusters “did not all correspond to predefined groups” (2382) but decline to discuss the non-correspondence further, in what might be seen as an overlooking of the genetic complexity of this region. A superficial reader might therefore overestimate the uniformity and distinctiveness of this region, as illustrated in Wade’s misreading, discussed below.

America and Oceania provide additional examples. Both include populations with considerable secondary cluster membership, with the American Pimas showing patches of purple amidst dominant green and the Oceanic Melanesians significant purple within dominant blue.²¹



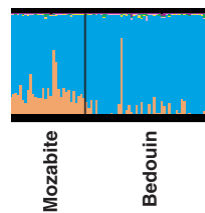
While the random ordering of lines in this case yields secondary patterns somewhat more coalesced than in the previous example, these are entirely overlooked in the discussion, where Rosenberg et al. write that in these regions, “inferred clusters corresponded closely to predefined populations” (2382). The secondary patterns I reference here are minor and scarcely derail the authors’ overall claims, particularly within Figure 2 as a whole, where most groups are indeed strikingly uniform. Still, describing both America and Oceania as having close correspondence between cluster and population encourages the reader to overlook the regions where

²⁰Individuals were placed according to their CEPH-assigned identification numbers (Cann et al. 2002). See Rosenberg (2007a, 2007b) and the researchers’ readme (<https://rosenberglab.stanford.edu/data/rosenbergEtAl2002/diversityreadme.txt>) and diversity (<https://rosenberglab.stanford.edu/data/rosenbergEtAl2002/diversitydata.stru>) files.

²¹The choice of purple coloration for these two groups was selected independently; the commonality does not signify genetic similarity across continents.

non-correspondence occurs, potentially predisposing an essentializing linkage between cluster and geography at the regional level.

Figure 1 generally fares better in this regard, given its greater overall regularity, but in certain areas population heterogeneity still appears greater than is addressed in the text. The authors write that “in the worldwide sample, individuals from the same predefined population nearly always shared similar membership coefficients in inferred clusters” (2382). Overall this is true, though the phrase “nearly always” implies somewhat greater similarity than the figures depict. In the case of Mozabite and Bedouin populations of the Middle East (Figure 1, $K=6$), both groups demonstrate significant orange membership in addition to the primary blue, with Mozabites uniformly reflecting dual cluster membership, as at $K=6$:



Bedouin membership in orange is more sparse, and this weaker secondary pattern is dispersed—as with the Biaka Pygmies—rather than aggregated. Each group thus appears dominantly blue, especially when the figure is zoomed out, reinforcing an impression of overall coherence for the blue cluster spanning Fig. 1. Such deviations are not discussed in the text.

That impression is strengthened in Table S2, which quantifies relative cluster membership for each region in the worldwide data. S2 registers a higher proportion of orange to blue for both groups, compared to other groups in the Middle Eastern region and beyond. Yet their distinctive properties are flattened in the “Middle East” line below, which averages out discrepancies between groups and establishes blue as the region’s dominant color.

Population	Orange	Blue	Yellow	Pink	Green	Purple
Mozabite	0.23	0.76				
Bedouin	0.06	0.93		0.01	0.01	
Druze		0.98	0.01	0.01		
Palestinian	0.02	0.95	0.01	0.01	0.01	
Middle East	0.06	0.92	0.01	0.01		

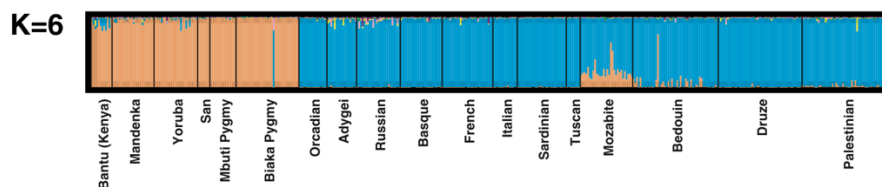
Given that other regions are even more strongly associated with their dominant clusters (Africa is 96% orange, Europe 97% blue), a hasty reader might take the table as confirmation of an essential parity between cluster and region, as an affirmation that the Middle East is blue and that its orange variations need not be regarded.

Admittedly, it would be impractical and methodologically unnecessary for the authors to address every deviation from the study’s overall findings in such a short paper, particularly as the results highlighted above do not detract substantively from their major conclusions. Yet the authors might have more openly acknowledged variations within the overall pattern of cluster membership and more assiduously avoided general statements of population uniformity. Such maneuvers would never fully deter racial interpreters but might have somewhat complicated their claims, which are indeed predicated on overstating the uniformity of clusters.

5.2 Worldwide Cluster Distinctiveness

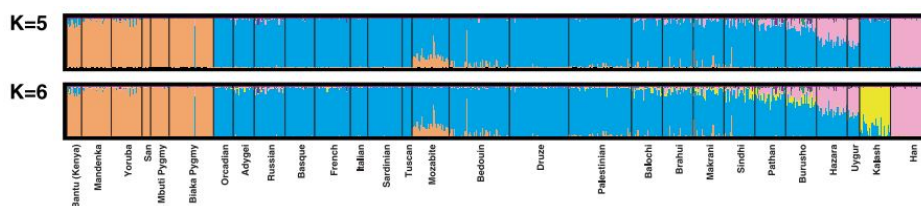
A second factor that may unintentionally precondition certain aspects of the racial interpretation is the ordering of population groups, which subtly foregrounds the appearance of distinct and bounded continental clusters. The populations in Figure 1 are arranged in roughly an east-moving direction from Africa through the Americas, presumably to highlight patterns associated with major geographic boundaries (oceans, mountains) and to reflect the approximate order of human colonization. The order of populations, unlike the ordering of individuals, is an aspect of figure construction determined by the authors, and in fact some continental and population groups are altered from their original order in the CEPH data (Rosenberg 2007b, 2). That reordering produces certain sharp divisions between clusters at the expense of softer, more graded boundaries.

For instance, the Mozabites and Bedouins, with their significant secondary orange coloration, are placed apart from the chiefly orange-colored African populations in the worldwide sample:



This distance results from Rosenberg et al.'s own reordering. Cann et al. designate the Mozabites as a North African population, placing the group directly after the other (Sub-Saharan) African groups and immediately before the three Middle Eastern groups. Following this original ordering would thus have placed the Mozabites, followed by the Bedouins, Druze, and Palestinians, immediately to the right of the other African groups. The effect would have been a clinal pattern, a slight gradient with the fringed orange of the Mozabites and Bedouins softening the boundary between orange and blue. Instead, by shifting the Mozabites to a Middle Eastern affiliation (eliminating the North African designation) and placing that entire regional grouping to the right of Europe, Figure 1 strands these groups in a sea of blue, diffusing the impact of their secondary clustering patterns and obscuring the genetic similarity between Middle Eastern and African groups. Intentionally or not, this implicitly posits the Sahara as divisive and encourages the perception of a sharp boundary on either side.

The placement of the Pakistani Kalash population is a second minor example. As the authors note, the group appears as its own, distinctively yellow, cluster at $K=6$. But at $K=5$ it is colored mostly blue, unlike the other Central/South Asian groups, which share some membership with the pink cluster dominant in the East Asian groups to the right:



The placement of the Kalash box at the far end of the Central/South Asia grouping therefore produces a sharp boundary between the two Asian clusters, interrupting what would otherwise appear as a fairly smooth blue-pink gradient at $K=5$. This placement derives from Rosenberg

et al.: Cann et al. locate the Kalash between the Burusho and Pathan groups, which are considerably more blue than the Uygur and Han groups neighboring the Kalash in Figure 1. Retaining the original order would therefore have preserved a smoother blue-pink gradient at both $K=5$ and $K=6$, with a lesser interruption of yellow at $K=6$.

In both cases, Rosenberg et al. decline to discuss the regional affiliations and placements of populations, an omission that may lead a reader to infer stronger genetic boundaries between the six clusters of Figure 1 than are actually present. This, in combination with the slight overstatement of population homogeneity discussed above, may predispose a reader to see the clusters as more uniform and more geographically determined than the data suggests. In the hands of Wade, and to some degree Sesardic, this becomes magnified into a full-fledged argument for geographical determinism.

5.3 *Rhetorical Intensification*

The racial interpretation of Rosenberg et al. (2002) describes the study as producing indisputable proof that its clusters adhere closely to sharply-defined regional and continental boundaries. In shifting toward a biological racial realist ontology, Wade in particular describes the inferred clusters as isolated, uniform, and geographically determined at both the population and continental levels.

At the population level, Wade represents groups as defined by single and unique clusters, particularly through exaggerating the article's discussion of linguistic isolation. The original authors posit language differences as one possible factor driving differentiation between groups: for instance, the Pakistani Burusho population, "a linguistic isolate, largely separated from other groups, although less clearly than the ... Kalash" (2383). Wade misrepresents both the degree and prevalence of this phenomenon, writing in generalizing terms that "Language is often an *isolating* mechanism that *deters* intermarriage. The Burusho, a people of Pakistan who speak a unique language, turn out also to be *unlike* their neighbors genetically" (98, emphasis mine). Placing a normative emphasis on isolation, deterrence, and dissimilarity, Wade removes the authors' hedges and falsely implies a total uniqueness for the Burusho population. In reality, Figure 2 reveals some overlap with neighboring clusters.

Moreover, Wade's use of the adverb "often" suggests that such isolation extends to other groups, an impression he reinforces with unwarranted extrapolation. The authors make no mention of linguistic isolation in Africa, but Wade writes confidently that:

Within races, the Rosenberg-Feldman study showed that different ethnicities could be recognized. Among Africans, it is easy to distinguish by their genomes the Yoruba of Nigeria, the San (a click-speaking people of southern Africa) and the Mbuti and Biaka pygmies. (98)

What Wade appears to mean by "easy to distinguish" is that these four groups are shaded four distinctive colors in Figure 2: orange, green, blue, and red, respectively. There are two problems with this representation. First, it isolates the Yoruba, whose orange cluster membership is nearly identical to their two (unmentioned) neighboring groups; it would thus *not* be possible to "distinguish by their genomes" the Yoruba from the Bantu or Mandenka. Second, Wade's commentary entirely overlooks the significant admixture of the Biaka Pygmy population as discussed in 5.1. In interpreting this group as colored solely red, he wrongly implies that these populations are characterized by unique and uniform clusters.

At the worldwide level, Wade persistently implies a particularly rigid separation between clusters, steadily minimizing discussion of clinal gradients in describing the five "main" clusters

as sharply, geographically bounded. He repeatedly emphasizes the fixedness of individuals and populations, writing, for instance, that the study “confirm[s] the remarkable extent to which people throughout history have lived and died in the place where they were born” (98)—a normative claim apparently of his own invention. From here, Wade further conceptualizes genetic flow as a discrete process of “splitting off” and “taking away” ancestral genes, phrasings that foreclose the possibility of clinal distribution or admixture by framing the movement of alleles across continents as a series of clean breaks rather than a continuous blending (98). Wade reemphasizes the separation of populations in discussing the two Asian regions, writing that “several Central Asian ethnicities ... are of mixed European and East Asian ancestry. This is not a surprise, given the frequent movement of peoples to and fro across Central Asia” (98). While he here admits the possibility of movement within continents, Wade refuses to describe this region as clinally distributed, even though its groups demonstrate notably graded, stable proportions of blue to pink. His phrasing rather emphasizes constancy of movement—a permanent flux—rather than stable, graded admixture; he wants to see this region as bounded by Asia on one side with Europe firmly on the other.

Wade thus attempts to use the study to buttress his views that individuals are sortable into discrete clusters determined by linguistics and geography, that they stay in place with minimal mixing, and that they adhere with few exceptions to continental boundaries. It is doubtful that Wade would have been dissuaded had the authors drawn more attention to the study’s placement of individuals and populations, as I have suggested. Nevertheless, his arguments here do appear to derive at least in part from the article’s slight and unaddressed overstatement of the uniformity and distinctiveness of clustering patterns.

Sesardic engages less directly with these aspects of the original article; however, his treatment of population genetics research in general forecloses the possibility of recognizing the data’s heterogeneity and clines in the first place. The section preceding his discussion of Rosenberg et al. (2002) is an extensive theoretical discussion of genetic classification. Detailing Edwards (2003) and others, it seeks to justify the validity of multi-variable approaches in sorting individuals into their populations of origin. Sesardic emphasizes the “virtual certainty” (151) of this process, which “approach[es] the limit of perfect accuracy” (151) and in which the wrong group categorization is “spectacularly unlikely” (152) given enough loci.²² The problem is that Sesardic uses this background to frame Rosenberg et al. (2002), falsely, as a comparable classification study designed to sort individuals into predefined racial categories. This impression is strengthened by his mention of the study’s unprecedented scale: the 1,056 individuals from 52 populations are the sole figures he furnishes, which after his theoretical discussion seem an argument for the “virtual certainty” of the results. The study, of course, was not designed to sort individuals into categories—it is of different design and focus than the projects described previously—but rather to detect a variable number of clusters as well as admixture. But Sesardic’s framing cleverly precludes the recognition of gradients or admixed individuals. By the logic developed immediately before his discussion of the study, these patterns would simply be viewed as artifacts or irregularities awaiting further resolution by yet more data. Indeed, his conflation of Rosenberg et al.’s methods with those of later studies assessing self-reported heritage, discussed in 6.2, furthers this impression.

Sesardic and Wade, clearly, do not fairly represent Rosenberg et al. in discussing the study. Their misappropriations are rather shaped far more strongly by their personal commitments to a biological racial realist ontology than by the article itself. As my analysis in this section has shown, however, both writers to some extent benefit from subtle overstatements of cluster

²²Sesardic echoes Edwards (2003) on the aggregation of loci as clarifying, but whereas Edwards’ reasoning works against racial claims (Kaplan and Winther 2014, 1043–6), Sesardic’s furthers the reverse.

uniformity and separation in the precursor article.

6 Correspondence

The final component of Rosenberg et al.'s racial interpretation is (iii): asserting a direct mapping between clusters and the five traditionally identified continental races, such that clusters become proxies for generalizations about racial phenotypes. This section shows how the original article relies on one particular, significant word choice that is directly seized by Wade and Sesardic in their unwarranted equation of clusters with races.

6.1 Correspondence

The word “correspond” is a touchpoint in Rosenberg et al. (2002), describing the relationship between clusters and geographical or linguistic groupings. It is closely associated with the study's main objective: the authors note that the project was designed “to test the correspondence of predefined groups with those inferred from individual multilocus genotypes” (2381). Though the authors describe varying degrees of correspondence in different aspects of their analysis, the word generally functions to affirm the methods and utility of the clustering algorithm in detecting meaningful population structure. While the term is used within an appropriate disciplinary context, I argue that its particular implementation in this article somewhat overstates the degree of the alignment between clusters and populations.

While the article often deploys this central verb with shades of meaning to describe relations of varying strength, such nuances may be easily overlooked given the strength of the word's associations. The authors occasionally seem to imply a one-to-one agreement, an equation rather than association between cluster and population. The verb does denote a certain degree of equality, signifying both a general similarity (“to agree with”; “to be congruous or in harmony with”) and a more precise physical alignment (“to answer or agree in regard to position, amount, etc.”) (Oxford English Dictionary). Whereas the authors thus may generally rely on the word in the first sense, the second and stronger definition is also connoted, an implication that may unintentionally deflect attention from individuals and populations that do not correspond perfectly to inferred clusters.

For instance, they write in summary that “genetic clusters often corresponded closely to predefined regional or population groups or to collections of geographically and linguistically similar populations” (2384). While the adverb “often” leaves room for some non-correspondence, it also indicates a particular frequency and firmness, particularly in combination with “closely” in this sentence. Likewise, the authors sometimes modify “correspond” with other strong intensifiers such as “largely” (2382) and “mostly” (2384). Stronger verb synonyms also appear, including *match* (used in the 2002 article and the 2005 follow-up study); and *anchor*, as in the observation that at $K=2$ “the clusters were anchored by Africa and America,” the verb calling to mind physical weight, heavy metal substances, and a rigid lack of motion (2382).²³ For a casual reader, these factors may combine with those discussed in previous sections to predispose an overlooking of complexity, leading to an assumption of greater geographical and population localization in the clusters than may actually be present.

In contrast, similar articles tend both to define “correspondence” as a metric and to note deviations from observed correspondence, as can be seen in two papers cited by Rosenberg et al. in support of their conclusions. Bowcock et al. (1994), which shows that microsatellite DNA

²³“Anchor” functions as a conceptual metaphor (Lakoff and Johnson 1980).

analysis allows the construction of evolutionary trees that “reflect [individuals’] geographic origin with remarkable accuracy” (455), does occasionally describe the alignment of genetics and origin using the verb “correspond,” but also uses more neutral verbs like “reflect,” “coincide,” and “tend.” Further, these authors forthrightly quantify the extent to which individuals deviate from the tree even though such deviations do not substantively detract from the main finding of “remarkable accuracy.” Similarly, Mountain and Cavalli-Sforza (1997) assess the “consistency” of an inferred tree corresponding to individuals’ population affiliations—in other words, they define a metric that in turn defines the observed correspondence (705). Throughout, they detail patterns of consistency and inconsistency, a discussion that (like Bowcock et al.’s) provides sufficient acknowledgement of difference while not impeding the main claim that “most individuals cluster with other members of their regional group” (705).

Rosenberg et al., unlike these precursors, rely almost exclusively on the word *correspond*, without defining or quantifying it, and decline to discuss deviations. They are, admittedly, working within both an established disciplinary vocabulary and a limited space. Yet the features discussed above may somewhat overstate the certainty of the findings, obscuring aspects of the data that do not reflect the overall pattern of correspondence and increasing the risk of the paper’s language being misappropriated—as happens, verbatim, in the racial interpretation.

6.2 *Rhetorical Intensification*

Both of Rosenberg et al.’s racial interpreters base their hereditarian claims on overstating the strength of correspondence between cluster and geography, emphasizing the strength of this relationship to imply that a variety of phenotypic “racial” differences must follow. Both rely heavily on the word “correspond,” using it disingenuously to conflate clusters with races.

Sesardic relies on the implied strength of the keyword “correspond” in mischaracterizing the study as a race-proving finding. He begins his population genetics section by attacking the concern that multivariate clustering analyses may produce distinct groupings that “do not correspond to common-sense races at all” (153), a warning further developed in a quote drawn from Pigliucci and Kaplan (2003): “While we argue that there likely are a variety of identifiable and biologically meaningful races, these will not correspond to folk racial categories” (1161). Sesardic’s framing, notably, thus twice establishes “correspondence” between clusters and races as the issue at stake and a connection he wishes to prove. This maneuver subtly conditions his discussion of Rosenberg et al. (2002), in which he emphasizes correspondence between clusters and *geography* in an effort to debunk the previously stated warnings:

Ironically, empirical knowledge about race and genetic[s] is advancing so fast that Pigliucci’s and Kaplan’s prediction was already refuted while the article with their bold claim was still in print. In an important paper that came out in *Science* at the very end of 2002, a group of geneticists showed that the analysis of multilocus genotypes of 1,056 individuals from 52 populations did allow an inference of group structure and that, furthermore, five clusters derived from that analysis of purely genetic similarities *corresponded* largely to major geographic regions (Rosenberg et al. 2002). This is an important discovery that makes it much more difficult than before to claim that race is entirely disconnected from genetics. (153, emphasis mine)

Despite claiming that this “important” study functions singlehandedly to “refute” the claim in question, Sesardic reports few conclusions, merely pointing toward the *correspondence* between clusters and geography. Yet he quickly concludes that these observations prove that race is

not “entirely disconnected from genetics,” a conclusion that seems to rest on little more than a misplaced idea of correspondence in presenting the relationship between cluster and geography as so tight that it confirms clusters-as-races in opposition to Pigliucci and Kaplan.

This careless substitution of one form of correspondence for another typifies a larger pattern in which Sesardic relies on “correspondence” to falsely imply that the study inquires into self-reported ancestry. His argument next turns to critics who have “attempt[ed] to downplay the importance of the results” by suggesting that its worldwide clusters “only loosely correspond to social categories of race” (153)—again establishing “correspondence” as the point of contention. Sesardic’s answer is Tang et al. (2005), a study of significantly different design that found a strong correlation between genetic clusters and self-identified race/ethnicity (SIRE) groups. Sesardic eagerly highlights the authors’ description of “near perfect *correspondence*” between clusters and SIRE and their observation that “the *correspondence* between genetic cluster and [SIRE] is remarkably high ... effectively synonymous” (Tang 2005, 268, 271, emphasis mine). Aside from the unremarked fact that SIRE is not necessarily synonymous with social categories of race as Sesardic implies, it is troubling that this elision of studies—as brokered by the repetition of “correspond”—is an attempt to misrepresent the later study as a refinement of Rosenberg et al.’s, implying that the earlier paper was always, in fact, about “race.” Sesardic essentially uses Tang et al.’s 99.9% correspondence to justify the designation of any non-corresponding data in the first study as irrelevant, resolved by the “near-perfect study” that followed. Again, Sesardic’s rhetoric depends crucially on an abuse of the strong sense of the word “correspond,” conflating geography and race without warrant. To be clear, there is little that Rosenberg et al. might have done to prevent this, although varying their word choice and more openly addressing areas of non-correspondence might have made Sesardic’s tricks more difficult, or at least more obvious.

In the larger context of Sesardic’s article, the suggestion that population genetics research nearly perfectly detects an individual’s racial and geographical origin grounds his ultimate gestures toward hereditarianism. Hochman (2016, 67) correctly notes that the article does not overtly make hereditarian claims, though it dangerously suggests them. Sesardic’s subsequent section discusses morphological racial characteristics, arguing that race is unmistakably inscribed on the body and that the multiplication of racial traits leads to an “*objective* biological classification” (156); the logical step linking this certainty to genetic determinism is implied, if not directly stated, by the previous section. Psychological differences, like IQ, Sesardic also wishes to see as non-arbitrarily related to genetics; again, following from his overblown discussion of population genetics, this would seem to leave little room for anything other than a hereditarian stance. So perfect is the match—the correspondence—between geography, genetics, and race that other attributes must surely follow.

Wade’s overt hereditarianism likewise depends on an inflated assessment of the “correspondence” described by Rosenberg et al. Like Sesardic, he champions the study’s finding that the worldwide clusters “correspond[ed] to the five continental races” (2014, 98), but in a formulation that eliminates the more cautious phrasings of the original. Elsewhere, he capitalizes on the authors’ keyword, intensifying it to suggest that in studies like this one, “everyone ends up in the cluster with which they share the most variation in common. These clusters always correspond to the five continental races” (96). His tone of absolute certainty implies a doubly perfect correspondence: of genetics to individual origin, and of clusters to races. While these implications are clearly exaggerated, they would appear to be enabled to some degree by the original authors’ language of matching and anchoring, extrapolated into a vision of inevitable, perfect alignment. “Correspond” seems an irresistible verb for a writer already prone to claims of geographical determinism.

Like Sesardic, Wade amplifies this implied cluster-race correspondence by mischaracterizing the study as continuous with others investigating self-reported ancestry. He introduces Rosenberg et al. (2002) by referencing an out-of-context excerpt from Risch et al. (2002), which states that numerous population genetics studies have “effectively ... recapitulated the classical definition of races based on continental ancestry” (3). That article in fact has nothing to say about social definitions of race, focusing rather on the utility of self-defined ancestry for biomedical applications. Yet Wade transitions immediately from here to Rosenberg et al. (2002), misleadingly framing it as “one of these more sophisticated studies” referenced by Risch et al. (who never discuss it). Borrowing the strong connotations of Risch et al.’s phrase thus enables Wade to imbue Rosenberg et al.’s mention of “correspondence” with a larger significance than it is accorded in the actual article. Wade’s sly insertion of a passing remark about “races based on continental ancestry” thus frames Rosenberg et al.’s observations about continental origin as having something to say about *race* rather than simply about genetic clusters. From here, Wade guides the reader through subsequent studies that are, in turn, falsely presented as validating the race-proving status of Rosenberg et al. (2002). His population genetics discussion ultimately descends, far more fully than Sesardic’s, to an elaboration of his hereditarian agenda. Wade’s hereditarianism depends centrally on a view of races as genetically inscribed and—as his misrepresentation of population genomics research suggests—detectible in the “correspondence” between cluster and geography, which is but a proxy for cluster and race.

I do not mean to imply that Wade’s many misleading statements follow entirely from Rosenberg et al.’s preference for the word “correspond.” Softer language and attention to areas of non-correspondence would certainly not deter such a motivated interpreter. His statements, however, do depend at least partly on the overwrought significance he attaches to the study’s use of “correspondence,” some of which is subtly suggested in the article itself.

7 Conclusion

This essay has endeavored to show how small details of “Genetic Structure of Human Populations”—the structuring of figures, particular phrasings—have been misappropriated and transformed in its racial interpretation, with Sesardic and Wade seizing on these to force a shift from the article’s bio-genomic cluster realism to their own biological racial realisms. My observations are not an accusation of carelessness, nor of any implicit racialized thinking on the part of the researchers. Rather, I suggest that these features arise chiefly from the constraints and conventions of scientific publishing. Rosenberg et al. (2002) is a slim three pages in length, an extremely compressed space in which to discuss the study’s limitations or exceptions to its major findings. Certainly, the verbal and visual features I have discussed are generally consistent with the article’s scholarly context and appropriate to its primary audience of other researchers. The authors, indeed, take pains to project neutrality in their writing. The article is mute on matters of race, impartially recorded, and is neither careless nor inflammatory in tone.

The problem of the article’s recurring racial interpretation stems more from its secondary, nonscientific audience. Writers like Sesardic and Wade are clearly irresponsible in their misuse of the paper, misrepresenting its methods and overstating its findings to advance their personal agendas. Still, it is worth asking how the article itself may have more fully prevented such interpretations from arising. There is of course a need for robust disavowals and correctives as to the paper’s scope, which many critics of Sesardic and Wade have undertaken. But it must also be recognized that other, non-specialist audiences exist, and that these are especially susceptible to misunderstandings arising from the sorts of subtle ambiguities I have detailed above. A series of small refinements might collectively have aided in the more accurate rendering of the article

in subsequent debates beyond the boundaries of science: drawing attention to aspects of figure construction, noting deviations from the dominant patterns, and defining key terms.

Admittedly, small refinements can only accomplish so much, and it is unlikely that the authors could ever fully block motivated writers like Sesardic and Wade. Nevertheless, it is incumbent upon the researchers—particularly in a field so loaded with pop-cultural significance as the genetics of race—to write with consideration of these secondary audiences. Faith in scientific objectivity can only carry so far in this context. Population genetics research, perhaps more than other fields, stands to be distorted and misrepresented in the service of arguments like the ones I have detailed in this essay. Excruciatingly precise language and a transparent disavowal of potential misinterpretations are necessary, even where scientists may deem these to be obvious or unworthy of accommodation. They are necessary because of the nature of this work, which is misreported because it is of high stakes to the wider culture: wittingly or not, it interfaces with some of our most entrenched and most dangerous assumptions about how people relate to one another, genetically, culturally, and politically.

Acknowledgments

I am grateful to Roberta Millstein and three anonymous reviewers for their insightful and constructive comments on earlier versions of this article.

Literature cited

- Bazerman, Charles. 1998. *Shaping Written Knowledge: The Genre and Activity of the Scientific Article*. Madison: University of Wisconsin Press.
- Bolnick, Deborah A. 2008. "Individual Ancestry Inference and the Reification of Race as a Biological Phenomenon." In *Revisiting Race in a Genomic Age*, edited by Barbara A. Koenig, Sandra Soo-Jin Lee, and Sarah S. Richardson, 70–85. New Brunswick: Rutgers University Press.
- Bowcock, A. M., A. Ruiz-Linares, J. Tomfohrde, E. Minch, J. R. Kidd and L. L. Cavalli-Sforza. 1994. High Resolution of Human Evolutionary Trees with Polymorphic Microsatellites. *Nature* 368: 455–457.
- Cambrosio, Alberto, Daniel Jacobi and Peter Keating. 2005. Arguing with Images: Pauling's Theory of Antibody Formation. *Representations* 89 (1): 94–130.
- Cann, Howard M. et al. 2002. A Human Genome Diversity Cell Line Panel. *Science* 296: 261–262.
- Coop, Graham et al. "Letters: 'A troublesome inheritance.'" Last updated on 8 August 2014. <http://www.nytimes.com/2014/08/10/books/review/letters-a-troublesome-inheritance.html/>.
- Duster, Troy. 2003. *Backdoor to Eugenics*. New York: Routledge.
- Edwards, A. W. F. 2003. Human Genetic Diversity: Lewontin's Fallacy. *BioEssays* 25 (8): 798–801.
- Evanno, G., S. Regnaut, and J. Goudet. 2005. Detecting the number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study. *Molecular Ecology* 14: 2611–2620.
- Fahnestock, Jeanne. 1986. Accommodating Science: The Rhetorical Life of Scientific Facts. *Written Communication* 3 (3): 275–296.
- Feldman, Marcus. 2014. Echoes of the Past: Hereditarianism and *A Troublesome Inheritance*. *PLoS Genetics* 10 (12): e1004817.
- Fujimura, Joan H., Deborah A. Bolnick, Ramya Rajagopalan, Jay S. Kaufman, Richard C. Lewontin, Troy Duster, Pilar Ossorio and Jonathan Marks. 2014. Clines without Classes: How to Make

- Sense of Human Variation. *Sociological Theory* 32 (3): 208–227.
- Gannett, Lisa and James R. Griesemer. 2004a. “The ABO blood groups: Mapping the history and geography of genes in *Homo sapiens*.” In *Classical Genetic Research and Its Legacy: The Mapping Cultures of Twentieth-Century Genetics*, edited by Jean-Paul Gaudillière and Hans-Jörg Rheinberger, 119–172. Abingdon: Routledge.
- Gannett, Lisa and James R. Griesemer. 2004b. “Classical Genetics and the Geography of Genes.” In *Classical Genetic Research and Its Legacy: The Mapping Cultures of Twentieth-Century Genetics*, edited by Jean-Paul Gaudillière and Hans-Jörg Rheinberger, 57–87. Abingdon: Routledge.
- Gross, Ari. 2012. Pictures and Pedagogy: The Role of Diagrams in Feynman’s Early Lectures. *Studies in History and Philosophy of Modern Physics* 43: 184–194.
- Gross, Alan G. and Joseph E. Harmon. 2014. *Science from Sight to Insight: How Scientists Illustrate Meaning*. Chicago: University of Chicago Press.
- Hochman, Adam. 2013. Racial Discrimination: How Not to Do It. *Studies in History and Philosophy of Biological and Biomedical Sciences* 44: 278–286.
- Hochman, Adam. 2016. Race: Deflate or Pop? *Studies in History and Philosophy of Biological and Biomedical Sciences* 57: 60–68.
- Jobling, Mark A. 2014. Trouble at the Races. *Investigative Genetics* 5: 14.
- Kaplan, Jonathan Michael and Rasmus Grønfeldt Winther. 2013. Prisoners of Abstraction? The Theory and Measure of Genetic Variation, and the Very Concept of “Race.” *Biological Theory* 7: 401–412.
- Kaplan, Jonathan Michael and Rasmus Grønfeldt Winther. 2014. Realism, Antirealism, and Conventionality about race. *Philosophy of Science* 81: 1039–1052.
- Lakoff, George and Mark Johnson. 1980. *Metaphors We Live By*. Chicago: University of Chicago Press.
- Mills, Charles W. 1988. “But What Are You Really? The Metaphysics of Race.” In *Blackness Visible: Essays on Philosophy and Race*, 41–66. Ithaca: Cornell University Press.
- Mountain, Joanna L. and L. Luca Cavalli-Sforza. 1997. Multilocus Genotypes, a Tree of Individuals, and Human Evolutionary History. *American Journal of Human Genetics* 61: 705–718.
- Myers, Greg. 1990. *Writing Biology: Texts in the Social Construction of Science*. Madison: University of Wisconsin Press.
- Orr, H. Allen. “Stretch Genes.” Last updated on 5 June 2014. www.nybooks.com/articles/2014/06/05/stretch-genes/
- Oxford English Dictionary. “Correspond.” Last updated 1893. <http://www.oed.com/view/Entry/41947>.
- Pigliucci, Massimo. 2013. What Are We to Make of the Concept of Race? Thoughts of a Philosopher-Scientist. *Studies in History and Philosophy of Biological and Biomedical Science* 44: 272–277.
- Pigliucci, Massimo and Jonathan Kaplan. 2003. On the Concept of Biological Race and its Applicability to Humans. *Philosophy of Science* 70: 1161–1171.
- Pritchard, Jonathan K., Matthew Stephens and Peter Donnelly. 2000. Inference of Population Structure using Multilocus Genotype Data. *Genetics* 155: 945–959.
- 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.

- 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: 2381–2385. <http://science.sciencemag.org/content/298/5602/2381>
- 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 Studies. *PLoS Genetics* 1 (6): 0660–0671.
- Rosenberg, Noah A. 2007a. “Distruct: A Program for the Graphical Display of Population Structure.” Last updated on 28 June 2007. <https://web.stanford.edu/group/rosenberglab/distruct.html>.
- Rosenberg, Noah A. 2007b. Distruct: A Program for the Graphical Display of Population Structure. *Molecular Ecology Notes* 4: 137–138.
- Serre, David and Svante Pääbo. 2004. Evidence for Gradients of Human Genetic Diversity Within and Among Continents. *Genome Research* 14: 1679–1685.
- Sesardic, Neven. 2010. Race: A Social Destruction of a Biological Concept. *Biology and Philosophy* 25 (2): 143–162.
- Sesardic, Neven. 2013. Confusions About Race: A New Installment. *Studies in History and Philosophy of Biological and Biomedical Sciences* 44: 287–293.
- Shiao, Jiannbin, Thomas Bode, Amber Beyer and Daniel Selvig. The Genomic Challenge to the Social Construction of Race. *Sociological Theory* 30 (2): 67–88.
- Tang, Hua et al. 2005. Genetic Structure, Self-Identified Race/Ethnicity, and Confounding in Case-Control Association Studies. *American Journal of Human Genetics* 76: 268–275.
- Taylor, Peter. 2011. Rehabilitating a Biological Notion of Race? A Response to Sesardic. *Biology and Philosophy* 26: 469–473.
- Taylor, Peter J. and Ann S. Blum. 1991. Ecosystems as Circuits: Diagrams and the Limits of Physical Analogies. *Biology and Philosophy* 6: 275–294.
- Templeton, Alan R. 2013. Biological Races in Humans. *Studies in History and Philosophy of Biological and Biomedical Sciences* 44: 262–271.
- Wade, Nicholas. 2014. *A Troublesome Inheritance: Genes, Race and Human History*. New York: The Penguin Press.
- Weiss, Kenneth M. and Stephanie M. Fullerton. 2005. Racing Around, Getting Nowhere. *Evolutionary Anthropology* 14: 165–169.
- Yudell, Michael. 2014. *Race Unmasked: Biology and Race in the Twentieth Century*. New York: Columbia University Press.

© 2017 Author(s)

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license, which permits anyone to download, copy, distribute, or display the full text without asking for permission, provided that the creator(s) are given full credit, no derivative works are created, and the work is not used for commercial purposes.

ISSN 2475-3025