# The Genetic Reification of "Race"?: A Story of Two Mathematical Methods

Rasmus Grønfeldt Winther

THE GENETIC REIFICATION OF "RACE"? A STORY OF TWO
MATHEMATICAL METHODS
*Rasmus Grønfeldt Winther, University of California, Santa Cruz*
*and University of Copenhagen*

**Abstract**

Two families of mathematical methods lie at the heart of investigating the hierarchical structure of genetic variation in *Homo sapiens: diversity partitioning*, which assesses genetic variation within and among predetermined groups, and *clustering analysis*, which simultaneously produces clusters and assigns individuals to these "unsupervised" cluster classifications. While mathematically consistent, these two methodologies are understood by many to ground diametrically opposed claims about the reality of human races. Moreover, modeling results are sensitive to assumptions such as preexisting theoretical commitments to certain linguistic, anthropological, and geographic human groups. Thus, models can be perniciously reified. That is, they can be conflated and confused with the world. This fact belies standard realist and antirealist interpretations of "race," and supports a pluralist conventionalist interpretation.

**Keywords:** mathematical methods; genetic classification; diversity partitioning; clustering analysis; reification

## 1. Introduction

Two mathematical methods lie at the heart of genetic classifications of human groups: *diversity partitioning* and *clustering analysis*. They are two sides of the same mathematics coin. Both are legitimate and consistent methodologies. As probability theory and statistics modeling practices, they cannot be questioned. This is the constructive part of the article.

Importantly, neither of these methodologies necessarily implies anything about the *reality* of human groups. This is meant in two ways. First, claims about the robustness of groups involve conventional choices based on assumptions about, for instance, how much genetic distance is required on average between groups for groups to be considered genuinely different and be granted ontic status. Different modeling assumptions yield different

outcomes. Second, the very data that are fed into the methods are neither unbiased nor theory-neutral. For instance, linguistic, archaeological and anthropological data are used to (pre-)define human groups; moreover, there may be visual or unconscious bias regarding which phenotypes of a particular group are sampled for genotyping.[1] Running models based on data taken from different sets of individuals of each population would yield distinct group diversity partitionings and cluster assignment of particular individuals, at least in principle if not in practice. Thus, model assumptions and interpretations, as well as data input, are subject to theory-ladenness and bias. This is the critique.

*The basic argument?* The modeling machinery may be well-oiled. The variables, functions, and derivations of diversity partitioning and clustering analysis are appropriately articulated and consistent, as we can expect from mathematical, formal systems. Yet model results are sensitive to (1) input, (2) particular assumptions made, and (3) output or interpretation. Reification of our expectations, biases, and preexisting theoretical maps (e.g., linguistic or geographic human groups) can occur in many places. In other words, we can make concrete things out of our abstract representations (i.e., reify) in various ways. That is, we might see or infer groups, natural kinds, and subdivisions when these do not exist. More broadly, arguments about the realism vs. reification of race are philosophical insofar as philosophy is critique (e.g., John Dewey, Michel Foucault). What is being critiqued in this article are the data and models used to answer empirical questions about whether our species is subdivided into groups (alternatively: natural kinds). As Bas van Fraassen recently pointed out to me (May 2014), these arguments are not philosophical in the broader sense of illuminating the aims (e.g., truth vs. empirical adequacy) and structure (e.g., syntactic vs. semantic) of scientific theory.

*The goal?* This article motivates the logic and procedure of these two methods and identifies places where reification can occur. The conclusion lists a set of open research questions that could inspire students of these topics.

## 2. A Story of Two Methods

What is the logic of these two families of methods? Very briefly, diversity partitioning assesses the amount of genetic differentiation present among distinct human groups. How much more genetically similar are two randomly chosen

individuals from the same group on average as compared to a randomly chosen individual from that group and an individual from another group (either from the same or a different continental region)? If there is no genetic differentiation among groups, then two individuals from the same group will, on average, be as different as two individuals from different groups. If there is some variation among groups, then two individuals from the same group will, on average, be more alike than two individuals from differing groups. The robust empirical result repeatedly found over the last forty years, using different molecular techniques on genes and proteins, and extensive global sampling, is that (approximately) 85 percent of all genetic variance is found within human subpopulations (e.g., Han Chinese or Sami), 10 percent across subpopulations within a continental region, and only 5 percent of the genetic variance is found across continents (i.e., "Negroid," "Caucasoid," and "Mongoloid"—Lewontin 1972 terms) (e.g., Lewontin 1972; Nei 1973; Barbujani, et al. 1997). No one questions these results, nor should they.[2]

In contrast, clustering analysis assigns particular individuals to groups (clusters) or to a specific weighting of more than one group. It also determines the gene allele frequencies of the clusters, under the assumption of a particular number of clusters. An individual is assigned probabilistically to the cluster or set of clusters that most closely matches what we call the "presence profile" of alleles in the individual. Intuitively, if an individual is *AA*, we would (ceteris paribus) prefer to assign it to a group with an *A* frequency of, say 90 percent rather than one with 50 percent *A*. But how do we assess the allele frequencies of groups? One way is to use Bayesian modeling. That is, we start with prior assumptions about gene frequencies in different populations (e.g., assume extremely high frequency of *A, B*, and *C* in some clusters, and extremely low frequency in others), and readjust the priors upon sampling of individual genotypes (e.g., Pritchard, Stephens, and Donnelly 2000). As long as a sufficiently high number of genetic loci are used (roughly 20 to 50), individuals can be assigned to clusters with extremely low probability of misclassification. The reliability of correct group assignation is very high. The computational power and programs for this task have been developed over the last ten years (e.g., Pritchard, Stephens, and Donnelly 2000; Edwards 2003; Rosenberg et al. 2002).

## 2.1. Diversity Partitioning

Diversity partitioning assesses how much of the total genetic diversity or variation in a species exists among individuals within predefined groups at

various hierarchical levels. That is, the amounts of genetic diversity found among individuals at the following three levels of groups are compared: (1) within single, local groups, (2) within different intra-continental groups, and (3) within different inter-continental groups. In order to assess the hierarchical structure of genetic variation, we need to develop measures thereof. Two broad types of properties are used to ground measures of diversity or variation: *group* and *individual* properties. Classic population genetic theory (e.g., Sewall Wright 1965, 1969; Cockerham 1969, 1973; and Lewontin 1972, 1974a) focuses on group properties, especially heterozygosity (see below). Other recent theoretical and empirical work uses detailed genetic sequence information to compare the genomes of individuals, within and across groups. Excoffier, Smouse, and Quattro (1992) developed a technique (Analysis of Molecular Variance, or AMOVA) that uses standard analysis of variance (ANOVA) theory to partition genetic variation at three hierarchical levels: individual, groups, and across groups of groups (e.g., continental regions) within a species (e.g., Barbujani, et al. 1997 used this method; on ANOVA, see Lewontin 1974b, Winther 2014a).

A common measure of variation used in these statistical methodologies is Wright's *F*-statistics. The basic method is to assess the levels of heterozygosity (e.g., the frequency of 2*Aa* rather than *AA* or *aa*, for a single biallelic locus) at different loci, in distinct groups. The more similar heterozygosity levels are to each other across groups, the more genetically similar the groups are.[3] *F*-statistics are a group property. One of Sewall Wright's *F*-statistics is as follows:

$$F_{ST} = \frac{H_T - \overline{H}_S}{H_T}$$

Where $H_T$ is the total heterozygosity of the entire population, and $\overline{H}_S$ is the heterozygosity within each subpopulation (group), averaged across all subpopulations. The fundamental idea here is to compare total heterozygosity to average subpopulation heterozygosity. Total heterozygosity is calculated by first averaging all the allelic frequencies, for different loci, across all groups, and then using those allele frequencies to determine the *expected* heterozygosity of the total population. Average subpopulation heterozygosity is calculated by taking the *actual* heterozygosity of each group and averaging those across all groups. In the extreme case where all groups are fixed for either one or the other allele of biallelic loci, the $F_{ST}$ measure will be at its maximum of 1, since the average group heterozygosity will be 0. Groups will be maximally different genetically. Conversely,

when groups have exactly the same levels of heterozygosity, for one or more biallelic loci, the averaged actual heterozygosity will be the same as the expected total heterozygosity ($H_T$) and $F_{ST}$ will be 0. Groups would be genetically identical. The Shannon information theoretic measure of variation used in Lewontin (1972)"strong[ly] resembl[es]" the heterozygosity measure, as Lewontin recognizes (388). For our purposes here, we can consider Lewontin to have employed a kind of $F_{ST}$ measure.

An example might help. Recall the mythical high school biology story of eye color genetics, in which two blue-eyed parents (each *bb*) always have a blue-eyed child, but two heterozygote brown-eyed parents (*Bb*), can have a child with either eye color. Setting aside the fact that eye color actually involves many genes—it is *polygenic*—this simple example helps motivate the meaning of heterozygosity, as well as the Hardy-Weinberg principle (HWP). Imagine a single population with thousands of only heterozygote brown-eyed parents (*Bb*). After a single generation of mating, their children will have genotype frequencies of very close to 25 percent *BB*, 50 percent *Bb*, and 25 percent *bb*; the corresponding allelic frequencies will be 50 percent *B* and 50 percent *b*. Barring the action of mutation and random genetic drift, or other evolutionary forces, genotypic and allelic frequencies will remain the same throughout future generations of mating. Now consider a separate second group with thousands of only blue-eyed parents. You already know what the allele frequency distribution will be for their offspring: 100 percent *aa*. In general, HWP states that percentages do not change after the initial bout of mating. Now, in this hypothetical case $\overline{H}_S$ is 0.25 because it is simply the average heterozygosity across the two groups, calculated from genotypes: (0 + 0.5)/2. In contrast, $H_T$ is the expected heterozygosity in a single large population with allele frequencies equal to the mean allele frequencies in the two subpopulations. In the first population, with both brown-eyed and blue-eyed individuals, these allelic frequencies are 0.5 and 0.5, while in the second they are 0 and 1, giving *pooled* allele frequencies of 0.25 and 0.75. $H_T$ thus turns out to be 2(0.25)(0.75) = 0.375. Thus, $F_{ST}$ is (0.375 - 0.25)/0.375 = 1/3. The example chosen is among the simplest possible, and is intended to make the basic logic of $F_{ST}$ explicit. Whenever allelic frequencies differ across two or more populations, the pooled allelic frequency average will be different than the average of the allelic frequencies of each individual group, and $F_{ST}$ will be non-zero. There is inter-group variation. Consider two final observations. If the first group had consisted of *BB* parents, there would be no heterozygosity within *either* group

(i.e., $\overline{H}_S = 0$), but the pooled heterozygosity would be 0.5. In this case, $F_{ST}$ would be 1, its maximum. Second, whenever $F_{ST}$ is non-zero, there will always be excess homozygosity (alternatively: deficient heterozygosity), per the Wahlund Effect.

Let us turn to the way $F$-statistics relate to diversity partitioning. $F_{ST}$ and the two other hierarchical inbreeding coefficients related to it—i.e., individual-to-total population, $F_{IT}$, and individual-to-group $F_{IS}$—are used to calculate hierarchical variance partitioning. In fact, with the three group levels considered—in diversity partitioning of *Homo sapiens*—i.e., intra-group, intra-continental groups, and inter-continental groups—$F_{ST}$ must be calculated hierarchically, twice. Setting these complications aside, there are clean, relatively simple, and well-documented relations among inbreeding coefficients and variance components.[4] Indeed, $F_{ST}$ is equal to the between-group variance, at a given level. The general result using these typical methods of diversity partitioning in humans is the (approximately) 85 percent/10 percent/5 percent variance measures mentioned above.

A simple table adapted from actual data in Lewontin (1972, 1974a)[5] helps further motivate intuitions (see table 1).

TABLE I **ALLELE FREQUENCIES OF TWO DISTINCT GENES AS USED IN LEWONTIN (1972) AND (1974A), ACROSS "RACES"**

| Gene | Alleles | Caucasoid | Negroid | Mongoloid |
|------|---------|-----------|---------|-----------|
| Duffy | Fy | 0.03 | 0.94 | 0.1 |
| | Fya | 0.42 | 0.06 | 0.9 |
| | Fyb | 0.56 | 0 | 0 |
| Auberger | Aua | 0.62 | 0.64 | |
| | Au | 0.38 | 0.36 | |

Note: Frequencies are rounded from four to two significant figures. Empty cells indicate lack of data. See esp. Lewontin 1974a, 153.

I choose the Duffy and Auberger genes of *Homo sapiens* from among the seventeen genes Lewontin used because they are instructive contrast cases. Duffy shows very high variation among the standard "races" (terms from Lewontin 1972), with one allele, Fy, being practically absent in Europeans but almost omnipresent in Africans. In contrast, Auberger exhibits very little interracial variation. Lewontin and others since him have discovered that most of our genes are like Auberger. This is another way of saying that

approximately 95 percent of genetic variation is found within races. On the other hand, if most of our genes had been like Duffy, $F_{ST}$ would indeed be significant, potentially approaching 1. In that case, most genetic variation would have been between races. Emphatically, genes similar to Duffy in frequency distribution are relatively rare in the human population.[6]

Which ontology is inferred from the (approximately) 85 percent/10 percent/5 percent result? Many interlocutors have argued that these results show that there is not very much genetic differentiation between groups defined on geographic, anthropological, or linguistic criteria. They reject these categories as "real," or at least as "biologically relevant." In particular, racial categories are *testable empirical hypotheses* that the data ultimately rejects. (Or so the interlocutors argue.) Diversity partitioning methodologies indicate that the abstraction of "race" is neither grounded in, nor justified by, genetic data.

### 2.2. Clustering Analysis

Given the assumption of particular individuals belonging to either a single group (cluster) or to a specific weighted combination of more than one group (when multiple population ancestry— i.e., admixture—is suspected), and supposing that there is a certain specific number $K$ of groups, how can individuals be assigned to their appropriate groups or weighted fractions of groups? Most basically, the presence profile of alleles in the individual is matched to the group or mix of groups that most closely matches it. Genetic structure across loci is used as information to infer cluster membership. "Structure" here does *not* mean (but it *can* mean, particularly for phenotypic "racial" characters; see section 5 below) that if an individual has allele $A$, she will also tend to have alleles $B$ and $C$ (alternatively: if she has a certain facial morphology, she will also tend to have a particular skin color and a specific hair type). After all, such correlation assumes significant linkage disequilibrium (i.e., statistical non-independence across loci) among the three loci, for which there is no guarantee, and which empirically is often not the case, certainly not for the neutral microsatellites, RFLPs, and SNPs that are used in many of these studies. Rather, structure here means that if the alleles of a sufficient number of loci in a given individual are identified, then we can classify that individual as belonging to a particular cluster with high probability.

A brief thought experiment might help motivate intuitions about the logic behind clustering analysis. Consider two groups $\Delta_1$ and $\Delta_2$, with

systematically different gene frequencies. For three biallelic loci, A, B, and C, respective frequencies of the dominant allele {*A, B, C*} are {0.9, 0.4, 0.49} for $\Delta_1$ and {.05, 0.7, and 0.5} for $\Delta_2$. We now actually have cross-loci information about the likelihood that an individual belongs to a certain cluster. Think about it: if I told you that an individual has haplotype *Abc*, what would you *bet* is her cluster membership? The answer is $\Delta_1$. After all, *A* is practically absent in $\Delta_2$ and *b* is significantly more likely in $\Delta_1$ than in $\Delta_2$; admittedly, whether the individual has *C* or *c* provides very little information. More generally, model-based statistical analysis (through either maximum likelihood or Bayesian statistical methods, e.g., Pritchard, Stephens, and Donnelly 2000; Rosenberg et al. 2002, 2005) tells us with which (high) probability an individual belongs to $\Delta_1$.[7] I leave it as an exercise to the informed reader to evaluate how much money you would bet on in this case. The point is that with sufficient cross-loci information about the haplotype of individuals, we can safely identify the clusters to which an individual belongs.

The point is actually a bit trickier because the population allele frequencies are often not actually known but are themselves estimated from the data. This may seem viciously circular or overly cumbersome. But it is neither. As long as we decide a priori on a fixed number of clusters, *K*, the mutual fitting of cluster allele frequencies and individual genotypes can be calculated in a straightforward manner. Pritchard, Stephens, and Donnelly (2000) do this through iterative sampling. This is their simplest algorithm, also built into the computer program *STRUCTURE* (947):

Step 1. *Sample $P^{(m)}$ from* Pr $(P|X, Z^{(m-1)})$.

Step 2. *Sample $Z^{(m)}$ from* Pr $(Z|X, P^{(m)})$.

Where *m* indicates the step, *P* is the random vector of population allele frequencies, *Z* is the random vector of populations of origin, and *X* is the random vector for individual genotypes. Pritchard, Stephens, and Donnelly write: "Informally, step 1 corresponds to estimating the allele frequencies for each population assuming that the population of origin of each individual is known; step 2 corresponds to estimating the population of origin of each individual, assuming that the population of allele frequencies are known" (2000, 947). In other words, we start with prior assumptions about genotype frequencies (e.g., for three biallelic loci, assume extremely high frequencies of *A*, *B*, and *C* in some clusters, and low frequencies in others—i.e., high frequencies of *a, b,* and *c*), place individuals in the clusters that best match their genotype frequencies, recalculate genotype frequencies for the

clusters and adjust priors accordingly, and repeat this entire computational process until we have clusters in which Hardy-Weinberg expectations hold (per locus), and linkage equilibrium (across loci) exists.

Which ontology is inferred from the fact that individuals can be reliably assigned to robust clusters? This fact could be taken as evidence for the strong reality of human groups, though few interlocutors have made this exact sort of statement, perhaps because of the potentially reactionary political repercussions such utterances may have or may imply (but see the recent book, Wade 2014). In discussing "Lewontin's Fallacy," Edwards (2003) claimed that the argument that the "division of *Homo sapiens* into these [racial] groups is not justified by the data" is fallacious because it "ignores the fact that most of the information that distinguishes populations is hidden in the correlation structure of the data and not simply in the variation of the individual factors" (798). He does not make ontologically strong pronouncements in this article, but his arguments are not inconsistent with a position stating that continental region classifications are real. The background argument here seems to be that significant inferential reliability of assigning individuals to clusters supports the reality of human group classifications. The results of diversity partitioning and clustering analysis pull in opposite ontological directions. Even so, their mathematics is mutually consistent, as we will now see.

### 3. Internal Methodological Consistency?

The two methods reviewed in section 2 are distinct ways of *characterizing the hierarchical structure of genetic variation*. The former assesses the hierarchical composition of genetic variance by exploring how similar groups are to one another, at a given level of the hierarchy. The latter assigns individuals to clusters, or finds groups, through the use of Bayesian modeling strategies or other methodologies (e.g., Principal Components Analysis). The two methods are used with the same overall aim of assessing the hierarchical genetic structure of (human) populations, but they answer different questions and use distinct methods to do so.

They are mutually consistent. Indeed, *differentiation among groups (as measured by* F-*statistics) is the logical outcome of similarity of individuals within groups (as found with the program* STRUCTURE*), and vice versa.* Thus, neither method is "wrong" or invalid, though each may be used inappropriately if employed to answer a question it was not designed to answer.[8] Diversity partitioning correctly indicates that there is very little genetic differentiation

among races, and just a little more among populations within a race. (N.b., there is extremely little variation among human beings in general. We are all basically identical across most of our genome.[9] Only the sequences that vary are here considered.) Even though the vast majority of genetic variation exists among individuals rather than across groups, clustering into groups can *still* be done, if we assume that a certain number of clusters exist. All you need is a little variation (read: non-identity) among the allelic frequencies in different populations, and among the distinct continental regions. Indeed, to take the argument to its extreme, if two groups or clusters of individuals had identical frequencies at 9,999 loci, but differed in frequency at just one locus, they would be different groups. As observed in the last paragraph of Rosenberg et al. (2002): "The challenge of genetic studies of human history is to use the small amount of genetic differentiation among populations to infer the history of human migrations" (2384). One could add: "and to infer the group memberships of any particular individual (under the assumptions of the statistical model)."

While our two mathematical methods are consistent, their aims, questions of interest, and basic assumptions are distinct. Diversity partitioning is particularly useful for evolutionary analyses of the opportunity for selection and random genetic drift in hierarchical populations. Clustering analysis can be used for making medical predictions (e.g., Burchard et al. 2003, Kumar et al. 2010). (But see the first item in the numbered list in the concluding section 6.) In fact, as Helen Longino put it to me recently (May 2014), discussion about "the metaphysics of race" is only interesting and important in so far as it connects with work in other areas such as biomedicine, forensics, and physical and cultural anthropology. Whether race is taken to exist or not—and in which sense (e.g., biological and/or social)—impacts the practices and commitments of these fields. Naturally, highly charged ethical, social, and political questions enter. Moreover, it is interesting that either method can be used for inferring migration patterns, because structured populations (with sufficiently high $F_{ST}$) and clusters (inferred populations in Hardy-Weinberg and linkage equilibrium) correspond, to an extent at least, to historical lineages. In summary, the two methods are consistent but rely on distinct assumptions and have different purposes.

### 4. Loci of Reification

The mathematical methods described above are just naked mathematics. Their logic is watertight, but they are highly dependent on the assumptions

used in their construction (see also Gelman 2008; Winther 2014a and references therein). In exploring loci or places where pernicious reification can occur, let us start with the data input stage of modeling. Which presuppositions are made about the homogeneity of sample sizes across groups, and how representative are the samples of the group (e.g., consider the difference between sampling Han Chinese with a population size of approximately 1.2 billion, and the roughly 150,000 Samis of northern Europe)? Are the sample data points independent of one another and homoscedastic (i.e., error variance in data samples is constant across loci and across clusters)? Is phenotypic appearance used as a conscious or unconscious sieve for which blood and tissue to sample? (If so, insofar as there is any correlation between genotype and phenotype, the data points are not independent.) With which cluster/group definitions do or should we start, either in collecting data and defining the groups to be tested in diversity partitioning, or in collecting data and setting the Bayesian priors in our cluster analysis? Do geography, archaeology, anthropology, and linguistics provide a priori information for genetical studies? These are all questions about the reliability of the *data input* to our models: who is sampled and which groups are presupposed? Defining the reification of race as conflating our theoretical expectations stemming from other fields such as anthropology and linguistics—as well as our phenotypic biases—with the (genetic) world, then reification can easily occur in this stage of modeling.

In each of these two mathematical methods, an irreducible theoretical element exists vis-à-vis defining populations. Indeed, groups are presupposed—from linguistic and anthropological data—or at least highly predetermined. In diversity partitioning, the starting point is the set of groups already identified by phenotypic, geographic, or cultural (e.g., linguistic) characteristics. In other words, there must be a set of properties, invariably correlated to culture, that *gives* the classification against which genetic variation is compared. In clustering analysis using the computer program *STRUCTURE*, assumptions stemming from linguistics and anthropology may be used to help set the strong Bayesian priors. The parameter space of multilocus information is too large, and clustering possibilities too massive, to *not* narrow down—guide—the possible clusters using the abstract maps of culturally defined groups. Indeed: "[The Bayesian approach] also eases the incorporation of various sorts of prior information that may be available, such as information about the geographic [or linguistic, anthropological, etc.?] sampling location of individuals" (Pritchard 2000, 947). Might there, though, be purely acultural priors in *STRUCTURE*? Perhaps,

but I suggest that unconscious (which individuals are actually sampled) and even practical (e.g., which genes are most easily sequenced) biases could be an unavoidable part of the modeling effort. A representative, independent, and random sample of each and every human population (whatever these may be, given the *clinal, gradating relation* of human variation) is an extremely challenging undertaking. Of course, these claims regarding reification and the conventionality of group definitions in diversity partitioning, and in the Bayesian priors of clustering analysis, require substantiation through examination of particular case studies. As a thought experiment, consider what modelers might do if they found a clustering that cut across culture. Would they accept it or might they tweak it by putting in more cultural or geographic priors? A number of sociology of science and philosophy of science research projects await.[10] The pernicious reification of our preexisting theoretical maps about groups and populations is difficult to avoid.[11]

Let us now turn to model *output*. Once we have our $F_{ST}$ measures and variance decompositions, how much difference is enough? Is the human $F_{ST}$ of roughly 0.15 sufficient for calling intra-continental and inter-continental groups distinct and attributing ontic status to them, or must we meet the typical boundary $F_{ST}$ of between 0.25 and 0.30 (see Templeton 1998)? Second, given that "the problem of inferring the number of clusters, $K$, present in a data set is notoriously difficult" and because the "posterior distribution can be peculiarly dependent on the modeling assumptions made" (Pritchard, Stephens, and Donnelly 2000, 949), it is unclear exactly how to interpret the reliability of the clustering for any particular $K$. Indeed, note that a $K$ of 3 or 5 (see Bamshad et al. 2003 and Rosenberg et al. 2002, respectively) is sometimes perceived to be a true and natural cut of human genetic variation, reflecting continental regions. But this may itself be a reification. After all, *STRUCTURE* identifies "multiple ways to divide the sampled individuals into $K$ clusters when $K > 6$ (Rosenberg et al. 2002). For example, in 10 replicates, *STRUCTURE* found 9 different ways to divide the sampled individuals into 14 clusters . . . (N. Rosenberg, pers. comm.)" (Bolnick 2008, 76). Thus, the apparent naturalness of $K = 3$ or $K = 5$ is actually a conventional choice about how to interpret the robustness of modeling results, rather than a mirror of nature.[12] The problem is worsened by the fact that for high $K$, there is not even a robust clustering. And, we *need* to assume a particular $K$ to do the iterative sampling mentioned in section 2.2 above. How to interpret the model output depends on conventional judgments: choose the appropriate cut-off point of $F_{ST}$ and choose the appropriate $K$.

Reification of our biases and our theoretical interpretations can occur in many places in this modeling process, especially in the input and output phases. The argument here is not that the two modeling methods lack merit or that the mathematics is wrong. But in certain places of the modeling stream, it is hard to know how to interpret input and output, or how to apply the modeling machinery. An *assumption archaeology* is necessary (Winther, under contract). Suppositions of various sorts (methodological, ontological, data-analytical, etc.) need to be stated clearly and self-reflectively. Questions and aims must be explicitly articulated and understood. Both for epistemic and ethical reasons, critical care is required.

### 5. Can There Be Phenotypic "Races" without Genetic "Races"?

In all of this, the question remains whether phenotypic "race" could exist even if genetic "race" is a reification. Regardless of the genetic facts, are there phenotypic races? Consider this passage from Feldman and Lewontin (2008):

> Using skin color, facial shape, and hair form, all obviously largely genetically determined . . . no one has any difficulty in differentiating between a random person taken from West Africa, from China, from Norway, or from the tropical rainforest of the Orinoco basin. With only a little more subtlety one can differentiate Amharic-speaking natives of Ethiopia from Zulus, Chinese from Japanese, and villagers of Andhra Pradesh from Afghanis by external morphology. (90–91)

As we saw above, Lewontin is hardly a realist about genetic race, yet he seems to be endorsing phenotypic race here. The measurement and metaphysical status of groups may have to be assessed separately at the genetic and phenotypic levels. That is, even if genetic race turns out to be a reification, phenotypic race could be a "human kind" (Hacking 1995), something about which I remain completely neutral. Moreover, mapping the causal links between these two levels is also a potential research project (e.g., phenotypic race could impact genetic—i.e., heritable—population structure via sexual selection, as Charles Darwin argued). These empirical, technical, and philosophical questions about phenotypic race are worth further consideration.

### 6. Whither Two Mathematical Methods and Genetic Reification?

The overarching research project of which this article is a small piece is the search for methods for identifying promises and dangers of scientific abstractions (Winther, under contract). I wish to find criteria and norms that will allow us to differentiate between the generative and productive use of our scientific abstract maps (such as cultural or linguistic groups), and the dangerous application—i.e., pernicious reification—of such maps. In other words, are particular groups of certain sorts *grounded* in patterns of genetic variation, or is an ontological interpretation of a clustering a biased and "viciously abstract"—to use William James's locution[13]—imposition on genetic data? This article's case study of abstraction and reification in science allows us to think about whether (1) biologists are actually measuring something significant in their clusters, (2) they are justified in their *kind-making* (see Hacking 1995, 2007), and (3) there is any knowledge/power (empirically and normatively) to their use of clusters in making predictions and formulating explanations of human evolution, capacity ascriptions (e.g., intelligence and athleticism), and disease proclivities.

   Why should we care? It seems obvious that the political and social stakes are high. Our very understanding of what it means to be human is under question, as are (our understandings of) human freedom, potential, and dignity. The main purpose of this text is to help clarify the two mathematical methods before turning to deeper questions of the normativity and metaphysical nature of potential natural kinds of *Homo sapiens*.

   Here are a few research projects that could be developed going forward:

1. *Classification and Function.* Do the genes we use in classifying human groups have any functional or mechanistic relevance? Our material and theoretical technologies do not yet allow us to assess the *causal* relevance of whatever genes do differ in systematic ways across at least some clusters. Presumably *some* of these genes are involved in the making of phenotypes that seem to differ across human groups. (This is a controversial claim.) But until we have found ways of describing actual genetic causal networks, the functional relevance of cluster-distinct genes (i.e., "private" alleles or genes that have extremely different frequencies in the two populations) remains unclear. Which sorts of empirical studies must be made to identify genetic causation? Can our classificatory investigations shed any light on mechanistic questions?

2. *The SMEO-P Account of Modeling.* In previous work (Winther 2006a, b), I provided a simplified and linear account of the modeling process. Modeling consists in setting up, mathematically manipulating, explaining, objectifying, and pluralizing. The SMEO-P account emphasizes the importance of ontological and methodological assumptions in modeling. I have only just begun to excavate the assumptions at play in the two mathematical methods of diversity partitioning and clustering analysis.

3. *The Role of Biologists in the "Race Debates."* Regardless of empirically robust modeling outcomes, and regardless of whichever definition or criteria of group reality biologists may have, social and political considerations may force scientists to keep a low profile about which sort of realism/constructivism/eliminativism interpretation to give (Haslanger 2008 discusses these interpretations). Put differently, is the history of racial discourse and of the biologization, objectification, and reification of racial categories so violent that everyone must engage in an enlightened dialogue that states not only the biological facts and methodologies, but also the historical, political, and social context? Should we even keep the categories of "race," ethnicity," and "population" in the face of historical baggage? Do our mathematical methodologies suggest that we can ascribe group reality at least sometimes, and with some justification? What can and should biologists add to the discussion of the reality/reification of "race"?

**NOTES**

While still a professor at UNAM (Universidad Nacional Autónoma de México), I was part of a "critical genomics" reading group led by Dr. Carlos López Beltrán. Discussions about the discourses and histories surrounding concepts of "race" and "mestizo," both within and without science, were lively, detailed, and insightful.

The book chapter seeded further work via three routes. First, upon sharing the book chapter, I received invitations to give lectures at Cambridge University, UC Berkeley, the University of Cape Town, and the University of Copenhagen. Second, shortly before the book chapter went to press, I sent it to Dr. Jonathan M. Kaplan. We exchanged several emails and he subsequently presented some of the book chapter's ideas, together with his own, at the Konrad Lorenz Institute in the summer of 2011. We decided to work together and I have greatly enjoyed our collaboration, which has resulted in three publications to date, in *Biological Theory*, *Philosophy of Science*, and *Theoria: A Journal of Social and Political Theory* (South Africa). Each of these articles has been written fully jointly and each was peer-reviewed. We look forward to further collaboration, perhaps including a book. Finally, the Lewontin-Edwards conundrum stimulated the writing of research grants, and I am now Principal Investigator for a trans-university research cluster (UC Berkeley, UC Davis, Stanford University, and UC Santa Cruz) on "Philosophy in a Multicultural Context," focusing during the 2013–14 academic year on "Genomics and Philosophy of Race," and institutionally rooted via the Institute for Humanities Research at UC Santa Cruz, http://ihr.ucsc.edu/portfolio/philosophy-in-a-multicultural-context/?id=15003.

Two workshops (Stanford University and UC Davis) and a big public conference (UC Santa Cruz) will result in at least a collection of articles. I look forward to further collaborations and many more learning experiences on the topics I first had the pleasure to address in this article.

In editing this translation, various stylistic and grammatical infelicities have been addressed. Examples, conceptual clarifications, and a few footnotes and bibliographic items have been added. Insightful comments by Doc Edge, Helen Longino, Mette

Smølz Skau, Bas van Fraassen, and an anonymous reviewer were addressed to the extent possible in this translation for *Critical Philosophy of Race*.

1. Since writing this piece, work surrounding implicit bias and the "implicit association test," IAT (e.g., Greenwald, McGhee, and Schwartz 1998), was pointed out to me. The relation to my argument is not that biologists acquiring data are "racist," associating different moral weights to different groups. Rather, the link to the IAT is that there may be unconscious mechanisms essentializing and singling out particular phenotypes from each human population as the appropriate exemplar of individual phenotype from which to draw blood samples for genotyping.

2. Importantly, neither Edwards (2003) nor Smouse, Spielman, and Park (1982) dispute these results. Indeed, the latter article's central argument is that "if one utilizes a multiple-locus approach, one will discover that human subspecific taxonomy is quite efficacious, even with the sort of marker loci alluded to above [i.e., the results of Lewontin 1972]" (Smouse, Spielman, and Park 1982, 445).

3. Lewontin (1972) provides four conditions that correctly describe the characteristics of *any* diversity measure: "(1) It should be a minimum (conveniently, 0) when there is only a single allele present so that the locus in question shows no variation. (2) For a fixed number of alleles, it should be maximum when all are equal in frequency—this corresponds to our intuitive notion that the diversity is much less, for a given number of alternative kinds, when one of the kinds is very rare. (3) The diversity ought to increase somehow as the number of different alleles in the population increases. Specifically, if all alleles are equally frequent, then a population with ten alleles is obviously more diverse in any ordinary sense than a population with two alleles. (4) The diversity measure ought to be a *convex function* of frequencies of alleles; that is, a collection of individuals made by pooling two populations ought always to be more diverse than the average of their separate diversities [the *Wahlund Effect*], unless the two populations are identical in *composition*" (388).

4. These were pioneered by Cockerham (1973); Holsinger and Weir (2009) is a recent instructive review of these relations.

5. Lewontin (1974a) attributes the data to Cavalli-Sforza and Bodmer (1971).

6. See Rosenberg (2011) for a synthetic review of state-of-the-art knowledge on patterns of human genetic variation.

7. Pritchard, Stephens, and Donnelly (2000) put it well: "Our main modeling assumptions are Hardy-Weinberg equilibrium within populations and complete linkage equilibrium between loci within populations. . . . Loosely speaking, the idea here is that the model accounts for the presence of Hardy-Weinberg or linkage disequilibrium by introducing population structure and attempts to find population groupings that (as far as possible) are not in disequilibrium. . . . *Under these assumptions each allele at each locus in each genotype is an independent draw from the appropriate frequency distribution, and this completely specifies the probability distribution* $\Pr(X|Z, P)$. . . . [Where] $X$ denote the genotypes of the sampled individuals, $Z$ denote the (unknown) population of origin of the individuals, and $P$ denote the (unknown) allele frequencies in all populations" (946, emphasis mine; sentence order slightly rearranged, as indicated with ellipses). Note that upon estimating the allele frequencies in actual populations, and sequencing individual genotypes,

we can use Pritchard et al.'s (2000) Bayesian clustering approach, embodied in the computer program *STRUCTURE*, to infer the population(s) $Z$ of origin of the individuals. N.b., this article has been cited 11,753 times according to Google Scholar, May 14, 2014; on the same day, Rosenberg et al. (2002) had been cited 1768 times.

8. See Feldman and Lewontin (2008), 89–90, for another way to make this general point.

9. Subsequently to this piece, this point was made in note 3 of Kaplan and Winther (2013), 404 and note 14 of Winther and Kaplan (2013), 75. See also Barbujani, Ghirotto, and Tassi (2013).

10. Having recently learned about actual genetic protocols used in, e.g., Noah Rosenberg's lab, it seems that some of my worries here were unwarranted. See also section 3 of Kaplan and Winther (forthcoming).

11. Weiss and Fullerton (2005), and Kaplan (2011) provide brief, useful discussions of these points.

12. Kalinowski (2010) points to another related problem with *STRUCTURE*, which can also lead to reifications of clusters: "STRUCTURE is also frequently used to identify the main genetic clusters within species. In this second type of analysis, individuals are assigned to clusters . . . but $K$ is deliberately set to be smaller than the actual number of populations. . . . The mathematical model used by STRUCTURE was designed for clustering individuals into Hardy-Weinberg/linkage equilibrium populations. It was not designed for clustering individuals into groups of populations, and may not work as its users intend when this is done" (1–2). Kalinowski's simulations show that when too few clusters are chosen, *STRUCTURE* pools individuals who would be pooled with other individuals if a higher $K$ were chosen. Again, this is not a problem with the probabilistic mathematics, but is a problem with the interpretation of nature that we impose from our modeling result.

13. See Winther (2014b).

**WORKS CITED**

Bamshad, M. J., S. Wooding, W. S. Watkins, C. T. Ostler, M. A. Batzer, and L. B. Jorde. 2003. "Human Population Genetic Structure and Inference of Group Membership." *American Journal of Human Genetics* 72: 578–89.

Barbujani, G., A. Magagni, E. Minch, and L. L. Cavalli-Sforza. 1997. "An Apportionment of Human DNA Diversity." *Proceedings of the National Academy of Sciences* 94: 4516–19.

Barbujani, G., S. Ghirotto, and F. Tassi. 2013. "Nine Things to Remember about Human Genome Diversity." *Tissue Antigens* 82: 155–64.

Bolnick, D. A. 2008. "Individual Ancestry Inference and the Reification of Race as a Biological Phenomenon." In *Revisiting Race in a Genomic Age*, ed. B. A. Koenig, S. S.-J. Lee, and S. S. Richardson, 70–85. New Brunswick, NJ: Rutgers University Press.

Burchard, E. G., et al. 2003. "The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice." *New England Journal of Medicine* 348: 1170–75.

Cavalli-Sforza, L. L., and W. F. Bodmer. 1971. *The Genetics of Human Populations*. San Francisco: Freeman.

Cavalli-Sforza, L. L., P. Menozzi, and A. Piazza. 1996. *The History and Geography of Human Genes* (abridged). Princeton, NJ: Princeton University Press.

Cockerham, C. C. 1969. "Variance of Gene Frequencies." *Evolution* 23: 72–84.

———. 1973. "Analyses of Gene Frequencies." *Genetics* 74: 679–700.

Edwards, A. W. F. 2003. "Human Genetic Diversity: Lewontin's Fallacy." *BioEssays* 25: 798–801.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. "Analysis of Molecular Variance Inferred from Metric Distances among DNA Haplotypes: Application to Human Mitochondrial DNA Restriction Data." *Genetics* 131: 479–91.

Feldman, M. W., and R. C. Lewontin. 2008. "Race, Ancestry, and Medicine." In *Revisiting Race in a Genomic Age*, ed. B. A. Koenig, S. S.-J. Lee, and S. S. Richardson, 89–101. New Brunswick, NJ: Rutgers University Press.

Gelman, A. 2008. "Variance, analysis of." In *The New Palgrave Dictionary of Economics*, ed. S. N. Durlauf and L. E. Blume. Basingstoke, Hampshire: Palgrave Macmillan. http://www.dictionaryofeconomics.com/article?id=pde2008_A000098.

Greenwald, A. G., D. E. McGhee, and J. L. K. Schwartz. 1988. "Measuring Individual Differences in Implicit Cognition: The Implicit Association Test." *Journal of Personality and Social Psychology* 74 (6): 1464–80.

Hacking, I. 1995. "The Looping Effect of Human Kinds." In *Causal Cognition: An Interdisciplinary Approach*, ed. D. Sperber et al., 351–83. Oxford: Oxford University Press.

———. 2007. "Natural Kinds: Rosy Dawn, Scholastic Twilight." *Royal Institute of Philosophy Supplement* 82, no. 61: 203–39.

Haslanger, S. 2008. "A Social Constructionist Analysis of Race." In *Revisiting Race in a Genomic Age*, ed. B. A. Koenig, S. S.-J. Lee, and S. S. Richardson, 56–69. New Brunswick, NJ: Rutgers University Press.

Holsinger, K. E., and B. S. Weir. 2009. "Genetics in Geographically Structured Populations: Defining, Estimating and Interpreting $F_{ST}$." *Nature Rev. Genetics*: 639–50.

Kalinowski, S. T. 2010. "The Computer Program STRUCTURE Does not Reliably Identify the Main Genetic Clusters within Species: Simulations and Implications for Human Population Structure. *Heredity* 1, no.8. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3183908/.

Kaplan, J. M. 2011. "'Race': What Biology Can Tell Us about a Social Construct." *Encyclopedia of the Life Sciences* (ELS). Chichester: John Wiley & Sons.

Kaplan, J. M., and R. G. Winther. 2013. "Prisoners of Abstraction? The Theory and Measure of Genetic Variation, and the Very Concept of 'Race.'" *Biological Theory* 7: 401–12.

———. Forthcoming. "Realism, Antirealism, and Conventionalism about Race." *Philosophy of Science*.

Kumar, R., et al. 2010. "Genetic Ancestry in Lung-Function Predictions." *New England Journal of Medicine* 363: 321–30.

Lewontin, R. C. 1972. Apportionment of Human Diversity. *Evolutionary Biology* 6: 381–98.

———. 1974a. *The Genetic Basis of Evolutionary Change.* New York: Columbia University Press.

———. 1974b. "Annotation: The Analysis of Variance and the Analysis of Causes." *American Journal of Human Genetics* 26: 400–11.

Nei, M. 1973. "Analysis of Gene Diversity in Subdivided Populations." *PNAS* 70: 3321–23.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. "Inference of Population Structure using Multilocus Genotype Data." *Genetics* 155: 945–59.

Rosenberg, N. A. 2011. "A Population-Genetic Perspective on the Similarities and Differences among Worldwide Human Populations." *Human Biology* 83: 659–84.

Rosenberg, N. A., S. Mahajan, S. Ramachandran, C. Zhao, J. K. Pritchard, and M. W. Feldman. 2005. "Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure." *PLoS Genetics* 1(6)e70: 660–71.

Rosenberg N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky, and M. A. Feldman. 2002. "Genetic Structure of Human Populations." *Science* 298: 2381–85.

Smouse, P. E., R. S. Spielman, and M. H. Park. 1982. "Multiple-Locus Allocation of Individuals to Groups as a Function of the Genetic Variation within and Differences among Human Populations." *The American Naturalist* 119, no. 4: 445–63.

Templeton, A. 1998. "Human Races: A Genetic and Evolutionary Perspective." *American Anthropologist* 100: 632–50.

Wade, N. 2014. *A Troublesome Inheritance. Genes, Race and Human History.* New York: Penguin.

Weiss, K. M., and S. M. Fullerton. 2005. "Racing Around, Getting Nowhere." *Evolutionary Anthropology* 14: 165–69.

Winther, R. G. 2006a. "Fisherian and Wrightian Perspectives in Evolutionary Genetics and Model-Mediated Imposition of Theoretical Assumptions." *Journal of Theoretical Biology* 240: 218–32.

———. 2006b. "On the Dangers of Making Scientific Models Ontologically Independent: Taking Richard Levins' Warnings Seriously." *Biology and Philosophy* 21: 703–24.

———. 2014a. "Determinism and Total Explanation in the Biological and Behavioral Sciences." *Encyclopedia of the Life Sciences.* http://philpapers.org/rec/WINDAT-4.

———. 2014b. "James and Dewey on Abstraction." *The Pluralist* 9, no. 2 (Summer 2014): 1–28.

———. Under Contract. *When Maps Become the World: Abstraction and Analogy in Philosophy of Science.* Chicago: University of Chicago Press.

Winther, R. G., and J. M. Kaplan. 2013. "Ontologies and Politics of Bio-Genomic 'Race.'" *Theoria: A Journal of Social and Political Theory* (South Africa) 60, no. 136: 54–80.

Wright, S. 1965. "The Interpretation of Population Structure by *F*-Statistics with Special Regards to Systems of Mating." *Evolution* 19: 395–420.

———. 1969. *Evolution and the Genetics of Populations. The Theory of Gene Frequencies, Vol 2.* Chicago: University of Chicago Press.