What Genes Are, and Why There Are No 'Genes For Race'

John Dupré

(In *Revisiting Race in a Genomic Age,* eds B. A. Koenig, S. Lee and S. Richardson, Rutgers University Press, 2008)

Talk of the genetic basis of race has resurfaced in the aftermath of spectacular progress in the development of genetic technologies, most especially technologies that provide genetic tests allegedly "for race" that coincide quite closely with self-reported racial identities. In a much discussed Op-Ed piece for the New York Times (March 14, 2005), Armand Leroi, an evolutionary biologist at Imperial College, argued that the classic claim by Richard Lewontin (1972) that human variation was overwhelmingly within rather than between races, so that traditional racial categories could be seen as socioeconomic constructs, was based on an elementary statistical error.[1] Whereas taking all genes separately, Lewontin's claim was true, the clustering together of genes characteristic of particular groups was able to show a distinctive genetic inheritance to traditional racial groups. These developments carry a significant danger of lending new respectability to controversial speculations about racial differences in such politically charged characteristics as IQ.

Unfortunately, the further these discussions move away from the technical contexts in which these genetic tests originate, the more misunderstandings appear. Other chapters in this volume (see Bolnick, Fullwiley) explain in detail the statistical procedures to which Leroi is referring, and the limitations of conclusions drawn from them. As I shall try to show in this chapter, insights from recent genomic science have helped clarify and highlight the ambiguities and misunderstandings that threaten incautious interpretations of genetic data and even of the

very concept of a gene. I shall also show how misunderstandings of these concepts can and do lead to spurious conclusions of the kind just outlined with regard to race and that, in fact, the kinds of tests just referred to do nothing to underwrite traditional racial categories.

The topic can be approached by considering the apparently quite straightforward claim that there are genes for race. This may seem banal and obvious; physical characteristics such as skin color presumably have a genetic basis. Far from being banal, the claim just mentioned is so difficult to interpret as to be close to unintelligible. This is because of the great difficulty of making clear sense of its main terms. The difficulty in defining race is familiar: although it is quite widely accepted by relevant experts that race is primarily a socially constructed concept, as the present volume documents there are still many who think of the concept as fundamentally biological. And whether it is a social or biological concept, there is no agreement as to how many races there are. Other contributors to this volume will explain many of the issues here. Difficulties with the concept of gene may be less familiar. The first task of this paper will be to explain why this concept is so problematic. This will make possible the differentiation of various interpretations of the claim under consideration about genes and race and examination of the plausibility and implications of each of them.

Are There Genes?

To explain the complexities of the various uses of the term ‘gene’, it will be helpful to provide a very condensed and somewhat Whiggish history. Genetics is generally thought of as starting with the work of the Austrian monk, Gregor Mendel. Mendel’s famous experiments in the 1860s involved crossing varieties of peas which, after generations of work by plant breeders, consistently produced plants with known phenotypes. For example, he interbred lines of peas with yellow and green colored seeds. The first generation produced by this crossing was found to

have uniformly yellow seeds. When these hybrid peas were crossed with one another, the second generation was found to have 75% of plants with yellow seeds and 25% with green. Similar results were claimed with other observable features.

The (mildly Whiggish) interpretation of these results goes as follows. Plants are assumed to contain factors that produce the observable traits. Each plant contains two such factors, one of which is passed on to the offspring. If we call the factor that produces yellow seeds Y and the other G, we assume that the true-breeding lines have two copies of the same factor, and we refer to the lines as YY and GG. The first generation hybrids will therefore have one factor of each kind, which we refer to as YG. The observation that the first generation hybrids are all yellow, hence that YG plants have yellow seed, is interpreted as showing that the Y factor dominates. In the second generation, assuming that parents are equally likely to pass on either of their factors, the plants will be divided between YYs, GGs, and YGs in the ration 1:1:2. Since only the quarter that are GG will produce green seeds, this explains the quantities found in the classic experiments.

Also famously, Mendel's results were largely ignored until they were taken up by several scientists independently at the beginning of the twentieth century. In 1909 the Danish biologist Wilhelm Johannsen named these factors genes. And in the first few decades of the nineteenth century a highly successful research program, now referred to as Mendelian genetics, greatly expanded empirical knowledge of the transmission of traits from organisms to their descendants. The most influential embodiment of this program was the work of Thomas Hunt Morgan and his students and collaborators on the fruit fly Drosophila melanogaster.

The most crucial thing to note about Mendelian genes, the objects of study in this episode of scientific history, is that they were causes of differences. No difference, no genes. In this strict

Mendelian sense, there are no genes for traits that are universal in a population. This is a concept suited to evolutionary theory, where selection can only work on differences, and one that remains prominent in medical genetics, since medicine is centrally concerned with deviations from the norm—and hence with genetic peculiarities that cause differences. In light of this general point we can easily see that the idea of genes for race is highly problematic.

Even supposing, for the sake of argument, that belonging to a particular race is a biological trait at all, it is certainly not the kind of trait that could be the subject of a Mendelian experiment. We could not, for example, examine the offspring of two black people, or a black person and a white person, and decide how many of them were black. Whatever the phenotypic criteria are for these categories, they are not immediately accessible to inspection. This simple observation, incidentally, is already enough to throw serious doubt on the idea that race might be a biological trait, but we will continue for the moment with the counterfactual assumption that it is. Minimally, the problem is that race, even if it were a biological trait, would be far too complex a trait. Perhaps there are Mendelian genes for dark skin, hair texture, the shape of facial features, and so on, but race is at least a matter of there being many such traits. Some people have some but not all of the relevant set of traits. There is quite certainly no gene that makes the difference between being black or white, even ignoring, for the moment, the fact that there is an important social aspect to many racial categorizations.

The simple preceding point is important because a lot of talk of genes is still firmly embedded in the tradition of Mendelian genetics, a tradition that first explicitly licenses the idea that genes are <u>for</u> a phenotypic trait, the trait to which they make a difference, and second almost inescapably suggests the erroneous inference that the trait is caused by the gene. Medical genetics, as just noted, is still very largely Mendelian, in that its traditional and continuing

central concern is with genes that make a specific difference, resulting in sometimes devastating pathologies. Medical genetics, it is true, is now moving rapidly toward a concern with predisposing genes, specific alleles that make it more probable that a particular pathology will develop. Good examples are the BRCA-1 and BRCA-2 genes, which strongly predispose women to developing breast cancer. But this hardly brings us nearer to a promising line for understanding genes for race. The idea of an allele that increases the probability of belonging to a particular racial group is a nonsensical idea. Many of the hypothesized genes for complex properties—intelligence, sexual orientation, violence, and so on—that appear regularly in the popular press raise a similar problem of spurious assimilation to traditions of Mendelian research on heredity, a further reason to emphasize the conceptual pitfalls. But to get to some slightly more plausible lines of thought, we should briefly move to some more recent history.

From the early stages of this program it was widely, but by no means universally, assumed that Mendelian factors, or genes, would eventually turn out to be specific material entities. Quite quickly, a consensus emerged that these were located on chromosomes—threadlike structures that were observable with the microscopes of the time. This consensus was reinforced as techniques were developed that enabled the order of the genes along the chromosomes to be ascertained, techniques which also made the hypothesis that genes were physical entities increasingly hard to resist.

A turning point in the attempt to convert hypothetical Mendelian genes into something solidly material was the determination of the structure of DNA by Watson, Crick, and others in 1953. The molecule had a number of features that seemed essential for anything that could be the bearer of Mendelian genes. The very long sequence of varying components making up a DNA molecule could be seen to have the information carrying capacity, by varying the order of its

nucleotide components, to specify the many traits for which there were genes. The double helical structure, allowing the possibility of separating into two strands, each of which could provide the template for a new double helix, provided a mechanism for the indefinite transmission of this information. And DNA was also a sufficiently stable molecule to maintain with some reliability the information it carried. It was naturally hoped that Mendelian genes would turn out to be specific sequences of nucleotides in the DNA that made up the chromosomes. When, a few years later, the "code" through which triplets of nucleotides "represented" specific amino acids was discovered, revealing the way in which sequences of DNA could provide information for the production of functional proteins, such a hope seemed to some even closer to realization.

However, the last fifty years of molecular genetics can also be seen as a gradual unraveling of this attractive vision. First of all, functional proteins correlate very poorly indeed with the phenotypic traits that are of interest to the student of whole organism inheritance. Even quite simple traits turn out to be the result of developmental processes involving many different protein products and much else besides, and for more complex traits the number of proteins involved might be hundreds or thousands. Hence a particular gene, conceived as the sequence of DNA coding for a particular protein, would typically have no very specific phenotypic upshot. Not only did traits turn out to have many genetic causes, but a particular gene would generally contribute to the development of many traits.[2]

More recently, the situation has proved to be far more complex still. First, the assumption that identifiable bits of DNA sequence are even "genes for" particular proteins has turned out not to be generally true. Alternative splicing of fragments of particular sequences, alternative reading frames, and post-transcriptional editing—some of the things that happen between the transcription of DNA and the formation of a final protein product—are among the processes the

discovery of which has led to a radically different view of the genome. The relationship between stretches of DNA and protein products is already many/many. Coding sequences in the genome are therefore better seen as resources that are used in diverse ways in a variety of molecular processes and that can be involved in the production of many different cellular molecules than as some kind of representation of even a molecular outcome, let alone a phenotypic one.

Moreover most of the genome doesn't code for anything. When it was still assumed that the function of the genome was to code for proteins, this non-coding sequence came to be known as "junk DNA." As a more complex view of the genome is emerging, it has become an increasingly more plausible project to look for different functions of this material. It is understood that much of the non-coding DNA is nevertheless transcribed into RNA, and the list of identified functions of these RNA molecules is growing rapidly. DNA sequences at other sites attach to various chemicals in the cell which in turn affect the rate of transcription at related loci. And it is plausible that even parts of the chromosomes that do not have specific chemical functions may have structural importance. The structural configuration of the chromosomes will affect, for instance, which parts are accessible to the transcription machinery. The assumption that the genome merely stores information is becoming untenable, and it now appears rather as an object in constant dynamic interaction with other constituents of the cell.

A problem that emerges from all this and that is exercising a growing number of philosophers of biology is whether any coherent interpretation of the concept "gene" can be recovered from these complexities. The first part of an answer is well captured by the proposal of Lenny Moss (2003) to distinguish two kinds of usage, which he calls genes-P and genes-D. Genes-P are related to Phenotypes, and the biological tradition of Preformation. They are most obviously the genes for this or that phenotypic trait found in the Mendelian tradition. Genes for

cystic fibrosis or Huntington's disease are genes-P, as are the red-eye or wingless genes in Drosophila. Mendelian genes retain important, but highly circumscribed, uses, but they are quite unsuited to general characterization of the genome. Genes-D are understood in relation to Development. Genes-D are defined not by their phenotypic outcome but by their molecular sequence. Though they are often referred to by means of a protein for which they "code," it is important to be aware that the sequence is fundamental. So, for instance, the N-CAM gene, named for the neural cell adhesion molecule for which it codes, can actually produce perhaps 100 different isoforms of this molecule in different tissues and at different developmental stages. Genes-D are the functional constituents of the genome as these constituents need to be distinguished in order to understand molecular function and, more specifically, the way genomes contribute to the differentiation of cell function in development. The philosophical issue here is whether there is any canonical division of the genome into genes-D, or whether, as I and a number of commentators suspect, this is just a name for any sequence of nucleotides that may, for a particular investigation, be of interest to a particular group of researchers.[3] But what is clear is that genes-D cannot in general be identified with relation to their outputs even at the level of functional proteins, let alone at the level of phenotypic traits.

Kinds of Genes

To get a better sense of the diversity of contexts and uses in which the term 'gene' appears, I will now summarize some of the more prominent uses of the term with a view to exploring what relevance, if any, they might have to our understanding of human race. I don't claim that this is a complete list of such uses or the only way in which this concept could be conceptually divided. I hope this list will, however, illustrate the diversity of uses across the

Mendelian/molecular and gene-D/gene-P divides, and consequently make it clear how hazardous it is to talk about genes, or the genetic, without a good deal of clarity about what is intended.

1. The hypothetical cause of a phenotypic difference. This is the original meaning of the term gene in classical, Mendelian, genetics, and the standard gene-P. I mentioned, as an example, the gene for cystic fibrosis.

2. The physical cause of a difference. There are, of course, physical features of the genome corresponding to traits with Mendelian inheritance patterns. These may be point mutations, deletions, insertions, inversions, duplications, and so on. They are of continuing interest mainly in the study of genomically based pathologies and also in the application of genomic knowledge to the improvement of techniques for plant and animal breeding. But they are not the objects of primary interest in the study of normal development. “The gene for cystic fibrosis” actually refers to a large number of possible mutations in a particular part of the genome, so the relation of this concept to the previous one is not straightforward correspondence.

3. The physical cause of a trait (the gene for X). Such a thing can only be assumed to exist at all in so far as there is a Mendelian trait, in which case it will be the kind of thing described in 2. For most X, there is no genomic feature or set of features that can be distinguished as the gene or genes for X. This is very likely the case for most of the Xs for which genes are regularly announced in the popular press, contributing to massive popular confusion on the general nature of genetics. The expression “gene for X” is, according to Moss, the canonical expression of the failure to distinguish genes-P from genes-D, and very often signals a conflation of these two concepts.

4. Quantitative trait loci. This category constitutes a technical qualification of the negative remark at the end of the last category. Breeders interested in, for example, leaner cattle or bigger cabbages, can locate particular genomic loci that have particular relevance to such features, and use this information to improve breeding programs. These are "genes for big cabbages" only in the sense that changes in these loci tend to affect the size of cabbages more than other loci. These loci may have countless effects, and their effects on cabbages may be dependent on interactions with numerous further genetic and environmental factors. Quantitative trait loci can be identified for breeding purposes through genetic markers (see below).

5. An open reading frame (ORF). This is the sense of "gene" intended, more or less, when we are told that there are only 24,000 genes in the human genome. It is a bit of sequence that is sometimes transcribed as a block into RNA. These transcripts are then subject to the processes of alternative splicing, editing, and so on that result in the much larger number of proteins, probably at least 10 times the number of ORFs, though still a very speculative quantity. This is a concept located firmly in genomic studies, and it is unlikely to be confused with a "gene for" a specific trait.

6. A functional part of the genome. This very loosely defined concept may well be the way the concept is heading in technical molecular biology. This, I take it, is the paradigmatic gene-D. I gave the example above of the N-CAM gene, which plays a central role in the production of a set of closely related proteins. These can coincide with the ORF, or be much smaller genomic elements.

7. An error in the genome. This is the molecularized Mendelian concept in much of medical genetics. It refers to any peculiarities of the genome with pathological

consequences. For example, the cause of a particular case of cystic fibrosis is a particular error, one of the set of mutations corresponding to the Mendelian gene.

8. A genetic marker. These will be discussed below. Whether they are properly referred to as genes is questionable, but this is the concept that underlies many recent claims about genetics and race. A genetic marker is a specific bit of sequence that need not have any function or correspond to any natural unit, but which is used to locate a part of the genome, generally because it is close to some functionally interesting part, for which it can serve as the marker.

Race and Genes

I have already noted that the first category on the list has no possible relevance to race and by implication nor has the second. Whatever races are, they are not Mendelian traits, traits that are present or absent in any individual and which are transmitted in specific ratios across generations. (3) deserves a little more discussion. If there are biological features that constitute belonging to a particular race (and I continue to assume this for purposes of the argument) then surely there is some set of factors that causes those features. In considering this suggestion, it is worth recalling that all or most biological kinds encompass a substantial amount of variation.[4] Hence if there are causes that make something a member of a kind, these causes are themselves likely to be diverse. Anyone who thinks that particular races are objective biological kinds must admit that they are variable kinds with diverse memberships.

It must be admitted that the variability of a trait does not in general prevent genetic analysis of the processes involved in its ontogeny. Examples are the complex diseases—diabetes, Alzheimer's, cardiovascular disease—that are currently undergoing this kind of investigation. Two further points distinguish such cases from the case of race, however. First, it is important to

distinguish the basis of correct function from the basis of dysfunction. Various blows on the head can disrupt proper brain function, but there is no correct blow on the head that explains normal brain development. In the case of diabetes, for example, part of the genetic project is distinguishing different classes of genetic failure which call for quite different therapeutic responses. None of these failures is in "the gene for correct blood sugar regulation." But second, and most importantly, the fundamental reply is that race is not a biological trait at all, it is a social classification. So there is no candidate subject for genetic explanation. In some cases this is self-evident, as in the category of Black defined by the one-drop rule in the U.S. More generally, this conclusion is entailed by the failure of races to constitute credible biological kinds. What I mean by this, and the reasons for it, should become clearer in the later stages of this paper.

Of course, even though races are not kinds, and even if race is not a biological trait, several phenotypic features strongly associated with conceptions of race are. Most obvious is skin color. Skin color is not, however, a Mendelian trait but, like height, the kind of continuously variable character associated with many genetic loci. If one were interested in breeding people with darker or lighter skins, one could probably discover QTLs that would facilitate a sophisticated breeding program of this kind. But particular alleles contributing to skin color are quite certainly not located exclusively in members of one (socially defined) race. Apart from the fact that people of different races do, of course, interbreed, this is evident in the continuous variability of skin color.[6] And, again of course, there are people who count socially as white who have darker skins than some people who count as black (a well-tanned person from Southern Europe, say, vs. an American with one grandparent of African descent).

These elementary observations about variability of race-indicative phenotypes and the regular interbreeding between people socially defined as belonging to different races is enough to show that there is no interesting work to be done in this area by genes-D. One could investigate the developmental processes by which skin color, say, is determined, and this would include investigation of various genes in the sense of (6), above (genes-D). It is almost certain that a variety of developmental processes might equally be found to lead to the same skin color, and certain that no such precise developmental sequence would be exactly correlated with any particular socially defined race.

Development, the more or less species typical physiological trajectory of an organism, depends on a great variety of factors. Though the genome is of course an essential such factor, what genomic resources are deployed at any point in the developmental cycle depends on many other factors of many kinds. Familiar metaphors for the genome—blueprint, recipe, program, and so on—suggesting that the genome alone determines the development of an organism are entirely misleading. The resources required for development are sometimes divided between the genetic and environmental, but within such a division environmental resources will range from parental care and social context to the set of extragenomic factors passed from mother to offspring in the egg cytoplasm. Any of these "environmental factors," up to and including the social, can affect the chemical and physical structure of the genome in ways that will contribute to the determination of which genomic resources are exploited by the developing organism. These complexities of development and of genomic function explain more deeply the point stressed earlier in this paper: that it is in general quite mistaken to think of bits of the genome having specific functions defined in terms of phenotypic outcomes. So, finally, there is no reason to expect a particular set of genomic features to provide a complete causal explanation of

a feature such as skin color. Like other features, we should expect skin color to be the final outcome of various possible developmental pathways, exploiting a range of genomic and other developmental resources.

Genetic Testing for Race

With this background we can address the concerns that have been raised by recent claims that race can reliably be discerned with genetic tests. First we need to relate the relevant categories of genetic test to the various gene concepts that have been distinguished earlier in this paper. These genetic tests depend on large numbers of genetic markers, and though these are genes-D, in the sense that they are specific bits of DNA sequence, there is no necessary, or even expected, connection to functional units of DNA. The tests in question are essentially similar to the technologies used in criminal forensic genetics and paternity testing.[8] Human genomes differ considerably in detail and one current successor project to the Human Genome Project is the cataloguing of single nucleotide polymorphisms, SNPs, particular points in the genome at which different individuals are found to have differing nucleotides. A particular SNP can provide the basis for a specific genetic marker. If one tests an individual for a sufficient number of common SNPs, one can find a profile that becomes more unique the larger number of SNPs one looks for.[9] An important thing to note about all these technologies for genetic profiling is that the variation studied is preferentially drawn from parts of the genome that do not have coding functions. This is for the simple reason that the less functionally critical the sequence is, the more variation will accumulate in it. Variation in sequence with important coding function is likely to have deleterious effects on the organism, so that integrity of sequence is maintained both by internal editing processes and, failing that, natural selection.

As well as being clear about the relevant concept of gene here, we need to consider how race is being conceived. The simple answer is that race is being identified with geographical ancestry. Certainly the concepts are related. Traditional broad racial categories were assumed to coincide with origins in the world's major continents, and more local racial concepts are often identical to concepts based on ancestry: African-American means, more or less, having ancestors from (West) Africa. In the U.S. "Black" is often understood as synonymous with "African-American," though in the U.K. the former term is used much more widely to include more or less anyone with naturally (i.e. not environmentally induced) dark skin, and can include South Asians or aboriginals from Australia or New Zealand.

The reason why SNPs or other indicators of genetic variation can track ancestry is clear enough. Since SNPs appear by random mutations and are passed on to offspring, they will, if they are fortunate, diffuse slowly around and away from the populations in which they originally occurred. SNPs will for a considerable time be most common in those areas where they have originated. Thus, and especially where populations are less mobile, particular populations of SNPs will tend to characterize particular geographical regions. We may finally consider the significance of recent reports (Rosenberg et al. 2002; Bamshad et al. 2003), apparently disturbing to some, that a genetic test can give results that very reliably predict whether Americans categorize themselves as African American, White (or European American), or Asian. Such tests indicate whether a person has ancestors who came from a particular geographic area in which particular SNPs originated. Thus the ability to distinguish, through these tests, those Americans who identify as African American shows that those who so identify tend to have more ancestors from a particular region, presumably in this case West Africa. This should hardly surprise anyone. It should be stressed, though, that all of these SNPs will be found in many people who

don't identify as black, since racial interbreeding will ensure that they are gradually spreading through the wider population.

If one is trying to specify an individual from among a generally interbreeding population, the desirability of tracking non-functional, or minimally functional parts of the genome is clear. The more functional the locus, the more selection will work to reduce variability. This argument is less straightforward in the case of testing for geographical ancestry. If populations have adapted to local conditions, then genes involved in such adaptations will be the most reliable indicators of that ancestry. Markers linked to such selected genes will be the ones that will tend to spread through the population. As a matter of fact it is unclear whether such functional genes are available.[10] The majority of loci actually applied are variations that occur in all populations and differ only in their frequencies within populations. The best correlation with West African ancestry in U.S. populations is exhibited by the so-called Duffy null gene, a variation that confers almost complete protection against the Malaria Plasmodium vivax. This gene occurs in a very large majority of sub-Saharan Africans, but is rare in Europeans. This degree of bifurcation is not currently known for any other locus. The extent to which the genetic variation characteristic of geographical locations is due to such adaptive histories is unknown, though for reasons that will be briefly discussed in a moment, much of this variation is likely to be a great deal more local than even the geographically restricted racial categories currently being considered. In what follows I shall mainly consider non-functional SNPs.

It should be noted that the same process of geographic origin of genetic mutations does explain the concentration of genetic disease in people of particular geographic origins. Many genetic diseases are consequences of simple point mutations in coding sequences that lead to a pathological defect in a functional protein. Echoing the preceding discussion, selective processes

such as linkage to a selected locus or heterozygote superiority (as in sickle cell disease) will most effectively spread the deleterious genes, though drift may also be a sufficient explanation. Such localized genetic diseases lend some support to the idea that racial identification may have a use in targeting of genetic medicine, though the weaknesses of such a strategy should also by now be clear. First, ancestry and race are not identical concepts, and only ancestry has any relevance to the incidence of genetic disease. Racial self-identification is at best a rough proxy for a specific ancestry. Second, as with the functionally neutral mutations that are typically sought in ancestry testing, deleterious mutations will not be confined to people who identify as having ancestors in the relevant geographic region of their origin. There is distinct danger that exaggerated correlations will be assumed, and disease will be overlooked in people who do not identify with the right racial proxy groups. We may hope that as genetic tests become rapidly cheaper, diagnosis by racial classification will be a very temporary transition toward more general disease screening or individual genetic testing.

Are There Biologically Distinguishable Human Kinds?

There is a long philosophical tradition of asking whether the categories into which we sort things are in some way given to us by nature or rather imposed on the world. Categories such as the chemical elements or, though more controversially, biological species, are often assumed to be given by nature and discovered by us, and these are often referred to as "natural kinds."[11] Pencils, penitentiaries, and philosophy professors, on the other hand, are clearly humanly created categories. A formulation of the question about race that has been partially resurrected by recent genetics, is whether human races are natural kinds (see Haslanger, this volume). I take the answer to be an unequivocal no. The procedures just discussed for testing for geographic ancestry in America are effective because the large majority of African Americans

have ancestors in a relatively specific geographic region. This does nothing whatever to support broader racial categories. As noted above, in the UK people are classified as black if they are non-white and hence experience discrimination. It includes people with Asian origins as well as those of African and Afro-Caribbean descent. It would also include (no doubt a very small number of) native Australians or New Zealanders. This is about as heterogeneous a group of humans in terms of origin, and therefore genetic profiles, as it would be possible to construct. In the United States the category is more narrowly circumscribed in terms of descent from black populations in West Africa, though to the extent that the one drop rule is taken seriously this would make the category even more heterogeneous in terms of origin and genetics than the British version. At any rate, no serious scientist thinks these categories, even if the American category is interpreted in terms of some greater predominance of African ancestors, have any biological grounding that could justify any claim to the status of a natural kind.

However, as previously remarked, the human population does have some geographic structure. Relatively isolated populations of humans, as with most species, make minor but specific evolutionary adaptations to their environments quite quickly. Skin color has been mentioned as one superficial characteristic which is notably fluid in human micro-evolutionary history. Following Kaplan and Pigliucci (2003) I have discussed elsewhere (2003, ch. 7) the relevance of local human "ecotypes" to discussions of race. The point is just that while there have been, and continue to be, numerous very local human types adapted to specific local conditions, this is a vastly finer-grained classification than any standard racial category. This phenomenon very possibly explains such observations as the dominance of Kenyans among international marathon runners as the consequence of local adaptation to a culture involving

extensive running at high altitude. (It has also been pointed out, though, that this tradition of success promotes a culture of aspiration in this particular direction.)

Broad racial categories, at any rate, comprise large numbers of ecotypes that are likely to differ in most respects of local adaptation. All equatorial peoples share dark skins and related adaptations to high temperature and strong solar irradiation, but there is little reason to suppose that they share any other adaptations not specifically responsive to climate. This is, of course, why claims such as a correlation between race and IQ are so biologically implausible. Though it is possible that local adaptation may promote subtle differences in cognitive skill sets, no good reason has been offered why these differences should be common to all ecotypes in low latitudes. Given that race broadly defined is a social kind with no interesting biological grounding it is overwhelmingly plausible that familiar social explanations—less educationally enriched environments, subtly culturally biased questions, and so on—will be more relevant to explaining prima facie data of this kind.

Why Does All This Matter?

It is sometimes remarked that it is misguided and even dangerous to engage in debates about the biological reality of race. By doing so, it is said, one is offering quite unnecessary hostages to fortune. What if races did turn out be biologically significant categories? We would still have no reason to discriminate against people because they fell into a different biological kind from our own. After all, male and female are indisputably significant biological categories, but this provides no justification for treating women (or men) as inferior. This is all true enough, and it is certainly important, if only because of the last point about sexual kinds, to insist that biological difference is no simple justification for social discrimination.

However, I think it is important to engage with the biological issue. First of all, there is no very serious hostage to fortune involved. We know enough about race to be quite confident that races will not turn out to be significant biological kinds, and it is at least worth explaining recent developments in genetics which are liable to be interpreted as underwriting biological interpretations of race.

Second, and more importantly, although we should not (of course) unjustly discriminate against people on the grounds of difference, real differences can and do provide reasons for different treatment. The political consequences of sexual difference are a much more complex issue than those of racial difference. Minimally, the fact that most women bear children is a reality that cannot simply be ignored in a just society. The problem, or one problem, is to make sure that this fact does not lead to systematic and unjust disadvantaging of women. In the case of race, by contrast, there is no such difference and therefore no such problem. If, as some racists may once have thought, black people were an evolutionary experiment somewhere on the step from apes to white people, there would be a real question as to what differences in treatment, if any, this justified (as, indeed, there is beginning to be a debate as to whether we are morally justified in treating apes in quite different ways from humans). But such a sharp distinction between human races is, needless to say, biological nonsense. Racial categories group together highly diverse groups of people on the basis of multiply evolved and trivial surface characteristics, and it would be miraculous if there turned out to be systematic biological differences dividing members of socially distinguished racial groups. So there is no question of what differences there should be in the treatment of people of different races: there should be none. The only question is the political one of how we move from racially divided societies practicing racial discrimination to a situation in which race ceases to be a concept of any interest

to anyone. Addressing biological misunderstandings doesn't do much to get us there, but it provides a small part of the necessary groundwork.

Conclusion

Contrary to some popular misunderstandings there are no "genes for" race in any of the various senses of the word "gene." There is lots of local variation within the human species, as there is for almost any widely distributed species, but as migration, easier travel, and so on make the species increasingly panmictic, this variation is likely to become ever more dispersed. This variation, moreover, provides no grounding whatever for the much coarser classifications that make up traditional racial categories, or indeed for any other comparable higher-level categories. The human species is an unusually genetically homogeneous one, and there are no important natural kinds distinguishable within it. As I have also discussed, genetic techniques make it possible to identify the geographic origins of some of the ancestors of individuals. But this reflects random and insignificant changes that occur in local human populations, or perhaps superficial adaptations to very local conditions, not the discrimination of significantly different kinds. Recent biology has confirmed the conviction of those who have long insisted that racial kinds were social kinds, and undermined any possible argument for placing these kinds in the realm of the biological. In its broadest and most common understanding, the concept of race remains little more than the reified residue of racism.[12]

Endnotes

[1] Leroi attributes the statistical insight to Edwards (2003). Robust replies can be found in Graves (2005).

[2] I should note that the many/many relations between genes and traits were not unfamiliar to geneticists in the first half of the twentieth century.

[3] See various discussions in Beurton, Falk, and Rheinberger, 2000. For empirical evidence that biologists do not have a clear consensus on the meaning of "gene" see Stotz and Griffiths (2004).

[4] Issues about biological kinds are discussed in several essays in Dupré 2002.

[5] Many theorists do not consider biological species to be kinds, but rather historically delimited individuals. I won't address this issue here, though it is quite clear that human races could not constitute individuals in the relevant sense.

[7] Actually the story is, as usual, more complicated and involves different ways in which the melanin-bearing melanosomes behave and are distributed within the epidermis, but the details are not important here. See, e.g., Thong et al. (2003).

[8] Individual genetic fingerprinting is most often based on measuring repeated sequences of variable length in non-coding DNA. This is a somewhat simpler technology, but the basic point is the same, namely an inventory of variable aspects of the genome. It is also worth noting that most tests of this kind use loci on either the Y-chromosome or the mitochondria, thereby restricting their relevance to only male or female ancestors. This has the rather striking consequence that claims to ancestry at, say, 10 generations in the past will actually be based on the genome of just one of the 1024 ancestors in that generation.

[9] Actual testing, both for ancestry and for individual identity, in fact uses a range of variable genomic features. I mention SNPs in large part because they are the easiest to explain. The differences between these features are not important for the present discussion.

[10] Some commercial providers of ancestry tests do claim to use mainly or entirely functional loci, and I don't wish to query (or endorse) their claims. One reason it is difficult to do so is that most of these loci are proprietary information. As noted in the text, I don't think that anything fundamental is at stake.

[11] For general discussion, see Dupré 2002, especially chapters one, two, and eight.

[12] The support of the Economic and Social Research Council (ESRC) is gratefully acknowledged. This work was part of the program of the ESRC Centre for Genomics in Society (Egenis). The chapter has benefited greatly from comments on an earlier draft by various colleagues in Egenis, especially Christine Hauskeller, Staffan Mueller-Wille, and Maureen O'Malley. I also received very helpful comments from Sarah Richardson. Finally, my understanding of the topic was much improved by attendance at the Authors' Conference in Stanford in January 2006.

References

Bamshad, M. J., Wooding, S., Watkins, W. S., Ostler, C. T., Batzer, M. A., Jorde, L. B. (2003). Human Population Genetic Structure and Inference of Group Membership. American Journal of Human Genetics, *72*, 578-589.

Dupré, J. (2002). Humans and Other Animals. Oxford: Oxford University Press.

Dupré, J. (2003). Darwin's Legacy: What Evolution Means Today. Oxford: Oxford University Press.

Edwards, A.W.F. (2003). Human Genetic Diversity: Lewontin's Fallacy. Bioessays, 25, 798-801.

Graves, J. L. (2005). What We Know and What We Don't Know: Human Genetic Variation and the Social Construction of Race. Posted April 25, 2005, from http://raceandgenomics.ssrc.org/Graves/pf/

Kaplan, J. and Massimo, P. (2003). On the Concept of Biological Race and its Applicability to Humans. Philosophy of Science (Supplement), 70(5), 1161-1172.

Lewontin, R.C. (2005). Confusions about race. Posted, April 20, 2005, from http://raceandgenomics.ssrc.org/Lewontin/pf/

Lewontin, R.C. (1972). The Apportionment of Human Diversity. Evolutionary Biology, 6, 381-98.

Moss, L.(2003). What Genes Can't Do. Cambridge: MIT Press.

Rosenberg, N.A., Pritchard, J. K., Weber, J. L., Cann, H. M., Kidd, K. K., Zhivotovsky, L. A., Feldman, M. W. (2002). Genetic Structure of Human Populations. Science, 298, 2981-2985.

Stotz, K. and Griffiths, P. E. (2004). Genes: Philosophical analyses put to the test. History and Philosophy of the Life Sciences, 26, 5-28.

Thong, H.-Y., Jee, S.-H., Sun, C.-C., Boissy, R.E. The Patterns of Melanosome Distribution in Keratinocytes of Human Skin as One Determining Factor of Skin Colour. British Journal of Dermatology, 149, 498-505.