UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Use of ‘Race’ as a Variable in Biomedical Research

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Philosophy (Science Studies)

by

Sophia Efstathiou

Committee in charge:

Professor Nancy Cartwright, Chair
Professor Craig Callender
Professor Gerald Doppelt
Professor Steven Epstein
Professor Arnold Gass
Professor Michael Hardimon

The Dissertation of Sophia Efstathiou is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Chair

University of California, San Diego

2009

DEDICATION

This dissertation is dedicated to Costas, Bookie and Jeannie.

EPIGRAPH

The pedigree of honey
Does not concern the bee;
A clover, any time, to him
Is aristocracy.

By Emily Dickinson

TABLE OF CONTENTS

Chapter 2 SYMPTOMS

LIST OF FIGURES

ACKNOWLEDGMENTS

Research for this dissertation was supported by an Andrew White Research Scholarship during 2007-2008, as well as grants from the Science Studies and Philosophy programs at UC San Diego. I am thankful for this support.

Research for chapter 2 and Appendix 2 was supported by a UC San Diego Senate research grant awarded to Michael Hardimon in the summer of 2004. Parts of chapter 3 were presented as papers at different conferences to the audiences of which I am most grateful.

Research for this dissertation was also supported by the Center for Philosophy of Natural and Social Science at the LSE where I have been a visiting student from Fall 2006 until now. The LSE was a second academic home for me in these years. I want to thank Damien Fennell in particular for his friendship and conversations on found science.

I am thankful for synchronizing with Zara Mirmalek at UCSD –our exchanges are most valuable to me. Likewise, sharing a living space with Marianthi Papalexandri-Alexandri's art and practice taught me about finding and founding art.

I benefitted greatly from studying with Craig Callender, Jerry Doppelt and Steven Epstein. I am thankful to each for their questions and guidance on the dissertation but also for their earlier instruction in seminars which gave this project initial shape and direction. Arnold Gass gave me an insider's look by bringing me to his practice at the V.A. hospital at San Diego which was invaluable to this project. I hope the thesis does not disappoint them.

It was Michael Hardimon's work on race and a seminar that he taught on philosophy of race that drew me to this topic. Michael has helped me immensely to streamline and justify my thinking and was patient with me even when I put up a fight, acting like a second supervisor to me.

Finally, Nancy Cartwright I cannot thank enough. She has helped me at every stage of this project and read drafts that I was too afraid to show to anyone. She could understand what I was saying sometimes before I did and she had the patience to let me work my way through this

project. This thesis may not reference her work directly; but this particular phrasing of the problem of a multi-use of “race” and so the possibility of thinking up some solution I owe to her.

VITA

1996-2000	Master of Physics, Joint Honors in Mathematics and Physics, University of Warwick, Class I
2002-2005	Teaching Assistant, University of California, San Diego
2006	Master of Arts, Philosophy, University of California, San Diego
Summer 2007	Instructor, Philosophy, University of California, San Diego
2006-2007	Research Assistant, CPNSS, London School of Economics and Political Science
2007-2008	Andrew White Dissertation Scholarship Recipient
2007-2008	Teaching Assistant, Philosophy, London School of Economics and Political Science
2008-2009	Research Assistant, CPNSS, London School of Economics and Political Science
2009	Doctor of Philosophy, Philosophy (Science Studies), University of California, San Diego

FIELDS OF STUDY

Major Field: Philosophy of Science and Science Studies

Philosophy of Science
Professors Craig Callender, Nancy Cartwright, Paul Churchland, Gerald Doppelt

Science Studies
Professors Steven Epstein, Andrew Lakoff, Martha Lampland, Sandra Mukerji, Steven Shapin, Susan Leigh Star, Naomi Oreskes

Philosophy
Professors Georgios Anagnostopoulos Richard Arneson, Michael Hardimon, Wayne Martin, Sam Rickless

ABSTRACT OF THE DISSERTATION

The Use of ‘Race’ as a Variable in Biomedical Research

by

Sophia Efstathiou

Doctor of Philosophy in Philosophy (Science Studies)

University of California, San Diego, 2009

Professor Nancy Cartwright, Chair

The use of ‘race’ as a variable in biomedical research is facilitated by embedding ordinary concepts of race in particular scientific domains. The dissertation articulates a process for how this can happen. The process has two parts: 1. Finding and 2. Founding a concept in a scientific context. The results of this process are called “found science” by analogy to found art.

Chapter 1 TOOLS draws distinctions between different race concepts following those of Michael Hardimon and Sally Haslanger. These distinctions are used to analyze a selection of the critical discourse on the use of race variables in biomedicine. Chapter 2 SYMPTOMS asks a ‘dummy’ question: “Should race be used to approximate medically interesting human genetic variation?” Answers to this question offered by Michael Root, Abdallah Daar and Peter Singer and Ian Hacking are analyzed. The analysis demonstrates that a. understandings of “race” vary, b. responses to normative questions vary in relation to these understandings and c. there is a pattern regarding what type of race concept is used in what context to argue for what normative claims. This suggests an underlying process at work.

How can context-specific normative demands be met by one and the same race concept?

They cannot.

Rather there is a process whereby an ordinary concept –even one as tainted as ‘race’– may come to fit a context of science. This process is defined by analogy to how common objects get to be art in Chapter 3 DIAGNOSIS. The case of “race”-usage in biomedical genetics is then analyzed as a case of “found science”. I apply the frame in two genetics studies (Rosenberg et al 2002 and Tang et al 2005) to show that ‘race’ as used in the context of these studies is a founded concept: it is an ordinary race concept founded in a genetics context and found to behave as a concept of ‘genetic ancestry’ would.

Appendices to the dissertation include A1 Background genetics knowledge, A2 Analysis of discussion articles in the biomedical literature, A3 OMB race/ethnicity categories as founded concepts in the context of demography and A4 a RECIPE for Found Science.

Introduction

The use of the term “race” has a loaded history. Race concepts have been used to express claims that were scientifically false and used to treat people in ways that were cruel. And all this is not in the past. ‘Race’ is a concept that is still multiply understood and multiply used, and still used to questionable ends.

The practical problem my dissertation tackles is this: Given this vague, “socially constructed”, “unscientific” ordinary concept, ‘race’, a concept which has led us astray morally and epistemologically, what do we do when we need to or want to study scientifically the phenomena that the term “race” can describe? Because the variable “race” and “race/ethnicity” does seem to describe important regularities in the current U.S. biomedical context; and this is the case both in epidemiology and in emerging fields of biomedical genomics.

Important health outcomes and health behaviors are regularly stratified by race/ethnicity variables in the United States (Epstein 2007, Root 2000, 2001, 2003). And though so-called “racial health disparities” are often put down to the unequal distribution of socioeconomic and educational resources across these groups, studies in population genomics and genetic epidemiology claim that interesting genetic and genomic features are also distributed in significantly different ways across race/ethnicity classes, in the United States (Risch et al. 2002, Rosenberg et al. 2002, 2005, Tang et al. 2005).

The philosophical problem I address is this seeming paradox: A concept that was historically and admittedly historically used to formulate, promote and legitimate oppressive ideologies, a concept that was historically used to formulate mistaken because they were typological- biological theories about human diversity, a concept that was operationalized and standardized in the 1970s in the U.S. census to measure civil rights violations and thought until recently to be biological nonsense, is, it seems, the same concept that promises to deliver wonderful, socially sensitized, innovative work in epidemiology and genomics.

But how can that be? How can concepts that are as "bad" as ordinary race concepts are - by scientific, moral and analytical standards- be able to deliver useful scientific results?

I propose that there is a process through which an ordinary –because ordinarily available– concept becomes embedded in a scientific context. This is a process that must happen repeatedly before the ordinary concept can become scientific and it is a process that happens according to the interests, settings and practices particular to the scientific domain where the ordinary concept is to be embedded. Through this process a concept is often tagged by the same ordinary term; but it need not. Through this process the concept is transfigured from some ordinary one, to one(s) extraordinary and scientific.

I describe this process by an analogy to how an ordinary object becomes "art" in the case of found art so I call its results "found science". The process involves two steps: 1. "Finding" a concept in a scientific context as ordinary and 2. "Founding" it there as scientific; i.e. articulating it in the terms and interests particular to the context in question. The process is complete once the concept can be found again in the scientific context, as scientific.

The process described is I think a general one and one that happens all the time in all sorts of everyday practices. So I suspect that the description I provide for it here will be useful independently of the case to which I apply it, useful in describing how other ordinary entities may become "scientific" or more generally how entities can become installed and functional in new symbolic practices. But all this is at best speculative at this stage.

What I focus on is how an ordinary race concept becomes scientifically articulated in biomedicine and articulated differently in different biomedical domains according to the specific ontologies, tools and needs of each domain.

This particular case is an important one to examine now because it calls for action. Biomedical genomics is an emerging biomedical field whose practices are being routinized and regulated as we speak. Race-specific biomedical genomics is the source of controversy and potentially questionable science.

Using specific examples I describe how the same word, “race”, is used to tag very different sorts of things and how tracking a –same- word can lead to unwarranted scientific inferences, when what the word describes in different cases will be [and will have to be] different things. The benefit of found science is that, besides clarifying some of the conceptual confusion, it describes why multiple articulation of the same concept should happen when we move from its ordinary to its scientific uses: it would be through the particular steps of the proposed process that an ordinary concept can become honed in, transfigured and transformed, in the right ways, that can answer our already specialized scientific questions.

The more practical side to this dissertation should not obscure the significance of the philosophical tool proposed. The notion of “transfiguration” and the process I use to describe it gives one answer to the question of why what seems to be an ordinary concept can function in extraordinary and scientific ways. It gives a frame for settling some questions about the measurement of ordinary concepts as it claims that they’re not what gets measured. Found science gives a reason for debating whether concepts embedded in scientific domains should continue being called by their ordinary name. And it can be used to frame concerns of methodological pluralists: If an articulation of an ordinary concept is embedded in some scientific domain more work will be needed to relate scientific results secured using this concept to problems expressed using the same word but which are (1) ordinary concepts or (2) concepts embedded in a different scientific domain.

All these proposed uses I make of the process in the context of current race-specific medical research. I argue that “race” can refer to very different concepts, concepts honed in to fit with particular discipline-specific interests that have little to do with what we think ordinary races are. I claim that these notions of race can be described using terms other than ‘race’ starting with a general distinction between “sociorace” and “biorace” notions of race; concepts that distinguish human groups on the basis of what a social context or respectively a biological context may take ‘race’ to be. And I conclude that in order to do certain normative work, i.e. to better track causes of health outcomes that are stratified by “race” and to settle ethical questions regarding these

categories' use, more research needs to go into each of both these types of concept of race within each of social and biological scientific domains; and interdisciplinary work is needed to reposition these articulations relative to each other and to enable joint action.

The Timeliness of this Discussion

In June 2005 the United States Federal Drug Administration (FDA) approved the marketing of a drug for heart failure called BiDil to self-identified African-Americans. Though there was no genetic explanation for why the drug would work differently in subjects of different races, the FDA reported it as a "step towards the promise of personalized medicine" framing the event in terms evoking personalized pharmacogenomics[1].

This event is almost paradoxical given the recent history of using race/ethnicity categories in biomedical research. Race/ethnicity classes used in the United States census were federally standardized in 1977. They were initially to be used to monitor civil rights violations –largely as a result of anti-racist movements of the 1970s. But though explicitly branded by the Office of Management and Budget (OMB) as "social-political constructs having no anthropological or scientific basis" these census classes have been codified into U.S. biomedical practice and progressively so since the 1980s (Epstein 2007).

U.S.-based research of the National Institutes of Health (NIH) uses race/ethnicity census classes; researchers are encouraged and in cases mandated to design research that tends to "minority health" and fulfills "inclusionary" criteria that are checked by Institutional Review Boards (IRBs) using race/ethnicity census categories[2].

Similar trends are being reported in U.K.-based biomedical research. Since 1995 the U.K. Department of Health (DH) has mandated National Health Service (NHS)-funded research to use census categories set up by the British Office of National Statistics (ONS). For example, a DH initiative of 2001 urges researchers to gather evidence that "reflects the diversity of the population" and checks recruited study populations along the same lines. It is reported that U.K.-

[1] http://www.fda.gov/bbs/topics/NEWS/2005/NEW01190.html (Last accessed on January 20, 2008)
[2] See Steven Epstein's examination of this process of codification, Epstein (2007).

based researchers are increasingly adopting ONS categories in "their sampling strategies and analytical designs" and more are projected to follow suit as a response to these demands[3].

Political and biomedical interests in catering to the health needs of ethnically diverse national and global populations coincide with economic interests in utilizing a global biotechnological industry. These interests find a shared object in the acquisition of *genetic and genomic* epidemiological knowledge.

Genetic knowledge is being secured by both national and international projects. India and Thailand have embarked on SNP-genotyping studies and Iceland and Mexico are two nations that have already set up national genotyping projects. Besides nation-specific work, genotyping projects are also becoming the result of multinational work processes. And a frame that may orient cross-national genomic work is that of "race". For example the U.K.-based *Human Genome Organisation* (HUGO) is an international organization established in 1989 by a group of scientists originally involved in the *Human Genome Project* (HGP) whose chapter HUGO Pacific began co-sponsoring in 2005 the *Pan-Asian SNP Initiative* aiming to study genetic diversity and similarity within "Asian" populations. This initiative is bringing together scientists from China, India, Indonesia, Japan, Korea, Malaysia, Nepal, the Philippines, Singapore, Thailand and Taiwan to share their data.

Not all this genetic information is medically interesting or linked to race. But there is definitely an interest in the question of what (type of) genetic variation links up to what disease phenotypes and how. Major U.S. health organizations have taken up research in this direction. The Center for Disease Control (CDC) now fosters under the National Office of Public Health Genomics a *Human Genome Epidemiology Network* (HuGENet) with an aim to assess the "impact of human genome variation on population health & how genetic information can be used to improve health & prevent disease"[4]. And research of the United States National Institutes of

[3] Ellison et al (2007)
[4] http://www.cdc.gov/genomics/hugenet/default.htm and http://www.cdc.gov/genomics/hugenet/global.htm [Last accessed 14/12/08]

General Medicine (NIGM) pursued under the Pharmacogenetics Research Network (PGRN)[5] is populating a database of genetic variation linked to variable drug response that uses 'race' as one of the variables that tags this information.

Links between race and genetics are still under research; but they ARE under scientific genetic research. The Human Genome Project devoted 3% of its annual budget to the relation between genetics and race[6] not as part of funding given to medicine but as part of "ethical, legal, and social issues (ELSI)" investigated. On this website we read the following:

> DNA studies do not indicate that separate classifiable subspecies (races) exist within modern humans. While different genes for physical traits such as skin and hair color can be identified between individuals, no consistent patterns of genes across the human genome exist to distinguish one race from another. There also is no genetic basis for divisions of human ethnicity. People who have lived in the same geographic region for many generations may have some alleles in common, but no allele will be found in all members of one population and in no members of any other[7].

Now this is very carefully phrased. Notice the claim is that there are no subspecies within modern humans –the nod is to ancient populations potentially being classifiable as subspecies. Further the reference is made to there being no patterns of genes across the human genome that distinguish races from each other; which does not imply there are no other types of genetic pattern to so distinguish races. And similarly, the reference relevant to the poor genetic cache of ethnicity is made via 'alleles' (expressions of a gene) and how these are not distinctive of particular populations –which again does not disbar other types of genetic markers as useful for distinguishing these groups.

On the International HapMap Project website we are warned that:

> the results of the Project could be misinterpreted to imply that constructs such as "race" are precise and highly meaningful biological categories. In fact, the information emerging from the Project is helping to demonstrate that *common*

[5] http://www.nigms.nih.gov/Initiatives/PGRN/ The PGRN member groups study variation in human genes relevant to pharmacokinetics (drug disposition) and pharmacodynamics (drug action), and the relationship of such variation to drug response phenotypes. A search for "race" in the resulting knowledge base, PharmGKB gave 82 hits more than half of which were published articles (www.pharmgkb.org/) [Last accessed 14/12/08].

[6] See http://www.ornl.gov/sci/techresources/Human_Genome/elsi/minorities.shtml [Last accessed: 14/12/08]

[7] See http://www.ornl.gov/sci/techresources/Human_Genome/elsi/minorities.shtml [Last accessed: 14/12/08]

> *ideas about race emerge largely from social and cultural interactions and are only loosely connected to biological ancestry*[8].

On the same site, the individuals whose DNA is sampled in HapMap are described as belonging to "populations with African, Asian, and European ancestry" –the populations being relevant on the basis of "geography", not 'race', but still coming from the household regions of the three basic 'colors'...

The Perspective of This Discussion

The case of "race"-usage in science is of special interest because it exemplifies a case of science-in-the-making; a case of science happening on shaky grounds, both in epistemological and ethical terms; a case of science whose terms are shifting in meaning and use; a real-life situation, like so many others we might experience.

This dissertation examines the "trajectory" of the word "race", in this current episode of scientific practice; that is, it looks at texts, arguments, using the term and assumes that the word corresponds to some concept whose meaning is geared to argue for interlocutors' theses. It attempts to articulate what concept(s) this word tags. But it doesn't stop at a conceptual analysis: it enquires how and why the concepts that this word tags are what they are, in science, and offers some general answer.

I am treating the word "race" like a physicist would tread a dyed fluid particle in a turbulent flow. The particular description offered for what happens to a concept the word might tag, in this case, is perhaps relevant for describing what happens in general when an ordinary concept is embedded in contexts of science. I don't purport to prove this very general claim! But I hope to show it is a reasonable claim to make.

---

[8] www.hapmap.org The sampled individuals are anonymous and their phenotype and medical history is unknown. However we know that the samples come from 1) Yoruba people of Ibadan, Nigeria, 2) Japan, in the Tokyo area, 3) Han Chinese individuals from Beijing and 4) U.S residents with northern and western European ancestry in Utah.

Allow me to illustrate how the general frame of found science is an innovative and possibly useful one by relating it to a problem that is well appreciated for its importance in philosophy of science: the problem of measurement.

An Alternative to Nominalism and Realism about Measurement

Science is almost synonymous to measurement. And yet –or perhaps because of this– defining what measurement is is no simple task. There is one question in particular which arises in the context of the scientific realism debate:

Does a measurement operation really measure what it purports to measure[9]?

For example, does a thermometer really measure heat[10]? Or does an IQ test really measure intelligence?

Considering answers on offer for this question in tandem with what answer found science proposes will help show exactly how the perspective of found science differs from available ones.

OK. So let us reconsider these questions.

Does a thermometer really measure heat? or

Does an IQ test really measure intelligence?

To answer these questions we need to consider if the particular measure approximates what we ordinarily think of as 'heat' or 'intelligence', but also whether this phenomenon exists. Is there such a quantity in the first place to be measured?

There are two general positions regarding the nature of measurement. One approach is to treat a measurement method as definitive of the ordinary concept in question. For example, IQ measurement is one definition for 'intelligence' and the indication of a thermometer is one definition of 'heat'.

The other approach is to assume that measurements are ways of finding out about quantities that can be identified independently of measurement. So, taking a temperature or an IQ

[9] cf. Chang and Cartwright (2006) for more detail on the problem of measurement

[10] I am saying 'heat' in reference to how Chang describes measurement practices of De Luc (Chang 2004, 60)

test would be (if correct) one way to find out about the quantities ‘heat’ and ‘intelligence’, which we know are so to speak “out there”.

These two broad positions can be called a nominalist and a realist position about measurement, respectively. And they come with variants. Under nominalist approaches to measurement there are more extreme ones like operationalism. This identifies the meaning of a concept with what operations are used to measure it –with the result that every type of measurement will come with (a definition of) its own concept. This is a safe way to make sure that every measurement operation is correct: it will by default always be measuring its own concept. Though the ease-of–mind afforded by operationalism is only conditional on our letting go of any questions about what the things we measure may be independent of our measurements; and calling such questions non-sensical really (what concepts would we have to express them in?). This by some counters the interest of science in synthesizing knowledge obtained by different measurements to know its objects.

Another nominalist position, though a more moderate one, is the view of conventionalism which posits that though a measure defines a concept, we can choose one of these concepts as “correct” as a matter of convention. So in that case a synthesis of various measures obtained for say the concept of intelligence has one way to operate –by using the (conventionally agreed upon) measure of it as a guide. Conventionalism however involves that convention be the epistemic authority on which our science operates –in contrast to the higher hopes of some that this authority be Nature. Both versions of operationalism then seem problematic.

So then what about realism? Realism, views measurement as an activity aimed at discovering what the true value of a quantity is, where the quantity is assumed to exist independently of how we measure it. But of course the question arises: on what grounds should this assumption rest? If our epistemic access to the world is effected via measurement, (a proposition that is especially compelling in the case of entities that would otherwise be unobservable), then how are we to know that we are not wrong? A naïve realist position which assumes that measurements correspond straightforwardly to our ordinary notions is hard to

uphold in the context of how science is practiced and the actuality of often seeming to rely on measurement methods for defining concepts in lack of any other epistemic access to the phenomena in question.

I am not going to claim that found science solves this problem! Perhaps it moves it closer to some solution. But the point is that both these views are different from the view that found science proposes. And this happens because within found science measurement is conceived as articulation. This is no simple reproduction of the given if such an activity is at all possible.

Found science posits that an ordinary concept may come to fit a scientific context if but only if it can be transfigured appropriately. That is, if it can be re-articulated to fit the ontological or metaphysical assumptions and interests of the target scientific context. It claims that ordinary concepts (often) just aren't scientific concepts though they can sometimes be transfigured into scientific concepts. This is a minimum condition for them getting to be 'proper enough' to be measured.

Found science invents new concepts, transfigured from the original that (if the process of transfiguration is done well enough) can be measured properly by different methods *as if* they were real independent of measurement. This doesn't make the invented notions real of course. But the point is that the ordinary concept would not be what gets measured at all once we measure a so-labeled "ordinary" concept in science.

Now, what results from this process need not be real; but it is not ordinary.

So found science shows that we need not be either nominalist or realist about the ordinary concepts. Once these concepts are founded in a science, the question can come up. In the process of founding an ordinary notion an operational definition for the founded notion could become available, but it need not. Founding a concept can be a matter of embedding it in theory via a set of assumptions or it can be a matter of otherwise articulating its scientific relevance. But found science suggests that measurements, even if one wanted to be an operationalist about the issue, would not be operational definitions of the ordinary concept but rather definitions of founded, scientific notions.

So, for example, were we to debate the correctness or not of different measures for say "inflation" across economic or social policy settings, we would have to understand ourselves as possibly debating about measures for different founded concepts. As such conflicts of this type can look for different grounds for settlement; first intra-theoretical ones –relevant to norms that are specific to what the founded concept is supposed to measure within its home context, and then, inter-theoretical or inter-disciplinary ones, deciding on what common grounds action regarding "inflation" might need to be taken, and seeing what perhaps new founding or other mapping might be able to recruit results obtained under the different founded notions in each domain and understand it in this new shared context relevant to any new requirements imposed therein.

With respect to a realist approach to measurement, again found science would suggest that what is being targeted as of interest to "measure" is not any ordinary concept but a new, transfigured one, perhaps a fictitious one, but also one that –should it be real –would also serve our interests to find out about, as real. As such, found science makes visible –not only the possibility of error with which we are all well familiar– but also why the creative element in scientific production that may lead to good or bad, real or unreal, but non-ordinary concepts to be thus "discovered" is so impressively successful when it is.

Where we had to decide between multiple measures for one ordinary concept –found science posits we have multiple non-ordinary, founded concepts to consider and then possibly multiple measures for each of these concepts!

Is this "proliferation" at all an improvement?

I think it is. First of all it is not any sort of proliferation, but rather a target specific articulation, a transfiguration that occurs according to expressed scientific norms and interests that results in new "scientific" concepts. But at the same time it is not a process which thereby reduces founded concepts to a measure.

I agree with Cartwright that scientific concepts should be textured enough and interconnected with other theoretical knowledge to admit of and justify multiple measurements.

So I would be wary of thinking them operationally reducible. Though found science does not come with a view on the issue.

So,

Does a thermometer measure heat?

Does an IQ test measure intelligence?

The answer found science gives is that the thermometer measures temperature –an invented concept (cf. Chang 2004). And as for intelligence –it depends who is asking. There are plausibly concepts of intelligence founded within different scientific or quasi-scientific contexts such as developmental psychology or cognitive science that an IQ test would measure and some that it might not.

The above discussion does not solve problems of measurement. But it shows that there is another way to consider these problems. One that is perhaps less ambitious regarding what is to be achieved by science (as scientific concepts cannot but be laden with theory and so always subject to revision) but perhaps a perspective from which *science* as much as Nature can take credit for its successes (and too responsibility for its failures).

The Structure of this Dissertation And Its Argument

I structure my dissertation as follows.

First, I introduce tools I use in the body of my argument (Chapter 1 TOOLS). This is a place where I state distinctions between *manifest* and *operative* and *target* notions following Sally Haslanger, distinctions between different concepts of race following Michael Hardimon, and where I define two second-order race concepts (*sociorace* and *biorace*) to describe concepts of ‘race as a social class’ and ‘race as a biological class’ respectively. These distinctions are used to analyze authors’ usage of the term “race” in Chapter 2. There is one more tool defined in Chapter 1 that is used in Chapter 3. This is the process of “finding” and “founding” a concept in a context of scientific use. So Chapter 1 states my tools of choice relative to already available ones.

I am thus in a position to fully articulate my thesis: race is multiply used in the literature on the use of race in medicine. But not haphazardly so; race concepts in use seem to work as either notions of race as a biological kind or of race as a social kind. This is often expressed as a divergence between authors' manifest and operative race concepts. But this plurality that might in ordinary discourse indicate a mistake or by some the "false consciousness" of the users here I think indicates a founded concept –one founded in a scientific discourse.

This is a thesis I begin to support in the second chapter of this dissertation (Chapter 2 SYMPTOMS).

Should "race" be used to approximate medically interesting human genetic variation?

I examine what I take to be typical possible responses to this normative question: those of Michael Root, Abdallah Daar and Peter Singer, and Ian Hacking. I examine a. what notions of 'race' are utilized by authors writing on the topic of race and b. how the selected notions accord with an answer to this question. These in sum are the results.

Michael Root answers no, while Abdallah Daar and Peter Singer answer yes. They each base their answers on considerations on what races *are*:

- Root thinks races are real social classes; they may pick out biological patterns but they are not biological categories. Race *cannot* track human genetic variation and it should not be used to approximate human genetic variation.
- Daar and Singer think that races are biological classes that approximate ancestral geographical populations. Races *can* track genetic variation and for epistemological, practical and ethical reasons they should be so used.

Hacking would answer yes and no; that is, he would take his answer to depend on the particular situation in question. His answer ultimately relies on how 'race' *behaves* as a variable in bioscientific and other research –i.e. it depends on what the founded concept is.

- Hacking thinks that races are superficial biological classes. Whether or not "race" should be used to approximate human genetic variation depends on what race can be *used for*: his account primes "statistical significance", "statistical meaningfulness" and "statistical

usefulness" as (further) normative grounds for (further) use. Race thus could but need not approximate biomedically interesting genetic variation.

The third chapter (DIAGNOSIS) answers the central question of this dissertation:

How can the concept of "race" secure all these answers?

Root's answer is coherent, if taken separately from Daar and Singer's, but it does not justify the fact that 'race' can be and is being used in practice irrespective of whether it *really* could or properly *should* be so used. Hacking's account gives no metaphysical story for why "statistical significance", "statistical meaningfulness" and "statistical usefulness" should be grounds for the use of "race".

I propose that we can use the word "race" to do specialized scientific work because the concept of race can be honed in to do different scientific work in different scientific fields.

So:

a. How is "race" used in biomedical research? Multiply, but not haphazardly.

b. What concept does "race" tag as used in biomedical research? Sociorace and biorace notions.

c. What are these concepts? "Race" tags a found/founded concept: which is really saying very little, unless you specify the context.

How *SHOULD* we use 'race'?

There is no general answer, only some (very vague) form for tabulating one…. This is a question I can formulate an answer to only after the *Found Science* account is made explicit, and more details are filled in such as costs/ benefits/ tools/ interests, etc. But the general direction is to collaborate. Founded notions will not suffice to address normative questions to a general satisfaction.

Chapter 1
TOOLS

Tools for Analyzing the Use of 'Race' as a Variable in Biomedical Research

"At the most general level, the task is to develop accounts of gender and race that will be effective tools in the fight against injustice."

(Haslanger 2000, 36)

The purpose of this chapter is to outline some of the tools that will be used in the rest of the dissertation. First, I distinguish between what I will call "manifest", "operative" and "target" concepts. Second, I present Michael Hardimon's definitions of particular concepts of race (the ordinary concept of race, the racialist concept of race, the populationist race concept) and one second-order race concept (socialrace). Third, I define two more general second order race notions, sociorace and biorace. Fourth, I describe what process I call "finding" and "founding" and explain how this would apply to describe what happens to a race concept when used as a risk factor in a biomedical setting. Finally I conclude with some clarifications regarding my usage of particular terms in this dissertation. The aim of this chapter is to define and familiarize the reader with some of the terms I am using throughout the dissertation.

1.1 Manifest, Operative and Target Concepts

I call *manifest* race notions the notions of race that authors explicitly define, or say they use. I call *operative* race notions the notions of race that authors actually use to argue for their theses, or the notions they would have to be using to argue for their theses. I call *target* race notions the notions authors explicitly propose should be used.

One could also distinguish between *manifest target* notions and *operative target* notions as the notions authors explicitly propose should be used are not always the notions needed to take the actions authors say should be taken. This would be then the manifest/operative distinction applied to the notions certain authors propose as target ones.

The authors whose work I am discussing here have not, to my knowledge, made the distinctions I am attributing to them. This dissertation reconstructs authors' work in terms of distinctions which are based on the work of Sally Haslanger and Michael Hardimon's.

1.1.1 How This Use of the "Manifest" v "Operative" v "Target" Distinction Diverges from Sally Haslanger's

The distinctions I draw are named after distinctions drawn by Sally Haslanger (2000), (2005a) who defines the concepts in her critical theoretical, or what she terms "ameliorative" framework. A brief description of how my use of distinctions between "manifest", "operative" and "target" concepts diverges from Haslanger's use (or of how my conception of what is manifest, operative and target diverges from hers) will make the framework of my analysis clearer.

Haslanger draws up these distinctions within a Marxian-inspired critical theoretical project. In such a project distinctions between what is visible or manifest and what is invisible and operative come with normative force. A divergence between the manifest and operative may indicate a possibly illegitimate force at work. This is not exactly the case in my project –one may think of force and violence in metaphorical terms, and this is in part what my notion of conceptual 'transfiguration' is based on, but admittedly what force ordinary concepts may be subject to in the process of becoming transfigured into scientific concepts is not here seen as illegitimate.

I will explain what I mean in more detail as follows. First I offer Haslanger's definitions. Then I compare the conditions under which she applies the terms, to the conditions under which I

apply the terms. Finally I conclude with an isolation of the relevant similarities and differences between our accounts.

1.1.1.1 Sally Haslanger's Distinction between "Manifest" v "Operative" v "Target" Concepts

Haslanger first distinguishes between what she calls a *manifest* concept, which is roughly "the more explicit, public and 'intuitive' one" from an *operative* concept, which is "the more implicit, hidden and yet practiced one" (Haslanger 2005, 14) in her work on "Ontology and Social Construction" Haslanger (1995). She later specifies some "axes of comparison" that the distinction between the manifest and the operative notion could map onto depending on what is relevant to describe. These are a. distinctions between institutional v. 'local' uses of a concept, b. public v. more idiosyncratic or individual uses, c. what is explicit in the minds of users v. what is implicit, d. what is thought v. what is practiced with a notion, or what we take ourselves to be doing v. what we're actually doing with it and e. what would be "appropriate" v. "inappropriate" uses of a concept (Haslanger 2005, 14).

These distinctions regarding the use(s) of a concept are drawn with a particular context of use in mind. Specifying a particular domain of use for a particular concept matters in deciding when the concept is a manifest or an operative one. I.e. an operative notion in context C could be a manifest notion in context C', given that contexts of use C and C' are different in the relevant ways –that is, in terms of an institutional v. local use of Xness, or a public v. private use, etc.

So, following Haslanger we might define when a concept would be manifest and when it would be operative, more generally, as follows (these are labeled 'Haslanger Manifest Concept', HMC, and 'Haslanger Operative Concept', HOC):

> (HMC) Call a concept of Xness *manifest* in context C, if and only if it is a concept of Xness that is used publicly in context C, or that is explicit in the minds of users in context C, or that captures what we think ourselves to be doing with Xness in context C, or whose use would be (taken to be) an appropriate one in context C.

(HOC) Call a concept of Xness *operative* in context C if and only if it is a concept of Xness that is used individually or more idiosyncratically in context C, or that is implicit in the minds of users in context C, or that captures what is practiced in context C, or whose use would be (taken to be) an inappropriate one in context C.

Beyond introducing the distinction between a manifest concept of Xness and an operative concept of Xness, Haslanger points to a third type of concept: what she calls a "target" concept. Haslanger defines what target concepts are operationally saying that "target" notions are what notions we decide, as a result of normative (later termed "ameliorative") enquiry, *should be* the notion had of Xness. In other words (Haslanger Target Concept, HTC):

(HTC) Call a concept of Xness a *target* concept if it is a concept that could effectively be used to further the goals of a normative (ameliorative) project.

Choosing what concept of Xness (whether it is an operative or a manifest or a new concept of Xness) we *should* be using in a particular context involves then defining this third type of concept: a *target* concept.

1.1.1.2 How Distinctions between "Manifest" v "Operative" v "Target" Concepts Work in This Dissertation

Recall my definitions of what notions are labeled "manifest", "operative" and "target" in this dissertation (1.1). There is one apparent difference. My use of Haslanger's terms could diverge from her own for the following reason: Haslanger defines these distinctions as applying to how groups of individuals, in a social context use concepts. I don't. That is, not explicitly.

I defined the notions as describing how authors use particular concepts in writing. But though the authors whose work I am examining are writing as individuals, they belong to and function according to norms of one or more disciplinary-specific, social contexts and I assume that –as they are publishing this work in disciplinary journals– it is from the perspectives, with the insights and responsibilities afforded to them in these roles that they write the texts I am analyzing here.

So when I analyze the work of Michael Root, I am analyzing the work of a philosopher who has been highly involved with philosophy of social science and epidemiology and is familiar with critical discourse in the social sciences. When I analyze the work of Abdallah Daar and Peter Singer I am examining perspectives of bioethicists, i.e. trained bioscientists who care for matters ethical. And in the case of Ian Hacking, I am analyzing the work of a philosopher of science with an expertise in the critical analysis of statistics and broader sociological and other sensibilities. These are then assumed to be no ordinary users of the word "race" and why observations I make in these instances provide grounds for me to posit what *found science* posits can happen to ordinary concepts so they come to be embedded in scientific contexts –though the analysis is of course not a proof of the claims of found science.

It is probably and unfortunately part of what limits of my philosophical analysis that it operates on such assumptions about "who" these authors are. I make these weaknesses visible to make them available to critical review. This is because I want to expose not only the "results" of what I am claiming in found science, but how I came to draw these inferences. What interests me here is not ordinary usage of the term race but rather analytical and scientific use of the term. But what is of added interest to me is also to engage a normative question of whether the use of race in science is what it should be.

I find that both of my concerns are addressed by examining discussions of the use of race in biomedical research. These are written by authors with an understanding of specialized usage – but with also an attention to the normative questions surrounding such usage.

So, I first analyze such a text to pick out a general form for the race concept used. This is because what matters in this context is form –a concern which I capture by defining the second-order race concepts 'sociorace' and 'biorace'. What matters is what *could* be being said and why it is this as opposed to another type of race concept that is chosen to say what *could* be being said in an analytical or scientific context. But I conclude that using these notions to satisfactorily answer normative questions needs more of an effort of synthesis, and an attention to other founded concepts, than can be achieved by possession of only one founded concept.

These steps I undertake here are steps of abstraction and inference that can easily get me into trouble: i.e. error. But these are risks I take as my aim –that follows the spirit of a critical project like Haslanger's– is not just to understand what people are saying –this should be more or less clear with a reading of their work– but also to understand why they are saying what they are saying, what concepts drive their arguments and also to understand why these and not other types of concepts are used, why these and not others *have to be* used in the context of these discussions given disciplinary backgrounds in question, and what notions *should* –at the end of the day– be used, to answer normative questions.

1.1.1.3 Divergences Between "Manifest" v "Operative" Concepts Need Not Indicate Illegitimate Oppression

My borrowing of terms from Haslanger's project is done with an understanding for the use she makes of these distinctions in her project; and with an interest in preserving some of the reverberations of the Marxian framework they evoke in the context of discussing conceptual transfigurations. A critical theoretical frame could be interesting to take up because such a frame is already functioning within a range of science studies accounts of scientific practice.

The possibility for such accounts is premised on an understanding of scientific practice or rather of theory-making as a form of consciousness. A form of consciousness is understood following Geuss (1981) as "a particular constellation of beliefs, attitudes, dispositions, etc" (12). As defined above a form of consciousness contains both discursive elements (elements with propositional content such as concepts, beliefs and ideas) and non-discursive ones (such as attitudes and dispositions which need not have propositional form). Thinking about discursive elements that make up scientific theories has perhaps had a longer history in the analytic tradition than thinking about the attitudes and dispositions of scientists. But including non-discursive elements in the subject domain of philosophy of science research was most notably achieved by Kuhn's philosophy of science and followed up by his colleagues and students.

Kuhn's work on scientific revolutions showed that philosophers of science could examine not only the "logic" of scientific discovery (as Hempel and Popper had) but also the social psychology and historical applications of scientific research. So since at least Kuhn's work the analytic tradition has described science in terms of both its manifest discursive, theoretical content as well as the values and attitudes of scientists (cf. work by feminist epistemologists like Helen Longino, Larry Laudan or Jerry Doppelt) and/or the values scientific work promotes or should promote via its function its historical social contexts (cf. for example Philip Kitcher).

Haslanger points to divergences between agents' manifest and operative concepts as symptomatic of agents' "false consciousness" and what may in their minds legitimate oppression that were they to be fully aware of their operative concepts they would dismiss or reject as unjustified (1995, 2000). She proposes that setting a (perhaps new) target notion can help emancipate agents from such conditions. Hers is a project following a Marxian-inspired, critical theoretical framework. Haslanger need not think that the inflicting of illegitimate oppression is always the case when manifest and operative concepts diverge (I am not sure that she does) but this is she thinks the case in the cases she finds interesting to look at when working on concepts such as gender and race.

Is the same thing happening in the case of manifest and operative notions of race in use in a biomedical scientific context?

It is not my starting assumption nor is it my conclusion that what is driving divergences between what race concept is manifest and what race concept is operative is always an interest in 'veiling' or 'oppression' –in the pejorative sense of these concepts, which is arguably the most common senses they are given. Divergences between manifest and operative –and target race concepts established in the context of this dissertation are in these cases found to indicate something different from what similarly labeled divergences to indicate in Haslanger's work.

My analysis also shows “manifest” and “operative” concepts as diverging –in the sense in which I defined these terms[11]. But this distinction doesn’t imply that what is going on here is illegitimate oppression (at least not conceived as oppression to agents within a social context). After examining what could have been the same sort of divergence in this instance, when manifest and operative notions of race seem to diverge, the conclusion is reached that if what phenomenon is illuminated here can have the character of an illusion and/or transfiguration a. it does so at a level more abstract than what Haslanger examines and b. with a purpose not all that contemptible –which, again, does not disbar a project pursued in critical theoretical vein, like Haslanger’s or those of science studies theorists, as interesting or useful. But which is I think a different project from the one of this dissertation.

It seems rather more enlightening, in reference to the cases discussed here, to view divergences between operative and manifest notions as symptomatic of notions being fit to and transfigured according to the particular ontological considerations of the scientific or other analytical context wherein they’re to be used, and as having been –and having to be– thus transfigured to serve possibly good, “ameliorative” ends, within these contexts. (No matter what use may thereby be made of these embedded concepts, within an actual, social setting –which can still, undoubtedly be invidious.) This is what process results in “found science”.

That this sort of ‘violence’ –inflicted on concepts– may accompany scientific practice is not new news. Heidegger ([1950] 2002) notices this and views it as a problem when arguing that science only sees as being what it can represent –and so do others. And some take the disconnect of ordinary from scientific contexts to be good too –for example Paul Churchland proposes that folk notions will be eliminated in time, by more precise scientific ones. What found science contributes to the discussion is the articulation of a process for going between these two types of context (an ordinary and a scientific one), which is demonstrated in the case it acts on an ordinary

[11] Whether this is also the case in the sense of HMC, HOC and HTC as indeed corresponding to agents operating within scientific social contexts is a matter of discipline-specific anthropological or social scientific empirical research needed to examine whether there is any value to the inferences I am summarily drawing here.

concept such as race, and how it becomes a founded notion, in the context of genetics (Section 3.3).

So, what I call manifest, operative and target race concepts are characterizations that have to do with the way a concept is used in an argument or text –but these terms are purposefully chosen to reverberate with discussions already in place in science studies and philosophy more generally.

There is another very useful tool I will be using: four race concepts specified by Michael Hardimon (2003), (ms1) and (ms2).

What could be the concepts of race that may be manifest or operative or target ones in the minds of "race" users?

This is the topic of the next few sections.

1.2 Hardimon on Concepts of Race

Michael Hardimon defines four race concepts i.e. four ways to articulate what we mean when we use the term "race". These are very useful definitions to make available for an analysis of how "race" is used in discussions about biomedical research.

I will be judging users' usage of "race" against these notions specified by Hardimon and calling the correlative concepts of race by Hardimon's terms where these accurately describe the notion in question.

Though Hardimon's concepts describe some of the race notions in use in these discussions, they cannot straightforwardly describe all of them. This is why I proceed in the next two sections of this chapter to define two vague race concepts which can do more of this descriptive work (section 1.3) and why I also describe a process, what I call "finding" concepts in one context of use and "founding" them in another, which this dissertation argues may be responsible for why usage of the term "race" comes out to be so systematized in this context (section 1.4).

1.2.1 The Ordinary, the Racialist and the Populationist Concepts of Race

Hardimon first carefully articulates what our ordinary concept of race consists in[12]. He does so by articulating the concept's "logical core" or its "intelligible nucleus" (Hardimon 2003, 441 and 442).

> Hardimon's *Ordinary Concept of Race* is our common sense concept of race that Hardimon proposes contains the logical core of all race concepts[13]. The logical core of the ordinary concept of race is specified in terms of three criteria that according to Hardimon all other race concepts that are "developments" of the ordinary concept will have to share. The following three theses "must be taken in conjunction" (442) and are all needed to come to understand the notion:
>
> 1. "The concept of race is the concept of a group of human beings distinguished from other human beings by visible physical features of the relevant kind" (442).
> 2. "The concept of race is the concept of a group of human beings whose members are linked by a common ancestry" (445).
> 3. "The concept of race is the concept of a group of human beings who originate from a distinctive geographic location" (447).
>
> I will refer to these as 'Hardimon Logical Core criteria' or 'HLC criteria' for short:
>
> HLC (1) visible *physical features* of the relevant kind (442)
>
> HLC (2) common *ancestry* (445)
>
> HLC (3) distinctive *geographic origin* (447).

The first criterion describes 'race' as a classification that sorts people into groups according *inter alia* to the way they look –in particular according to skin-color, facial features, hair texture, etc. In Hardimon's words "This thesis captures the basic intuition that race is essentially manifest" (442). The second criterion captures the idea that 'race' is inherited from

---

[12] An account of all four of Hardimon's notions is in "On the Ontology of Race", Hardimon (ms1), though Hardimon (2003) focuses on "The Ordinary Concept of Race" and Hardimon (ms2) on "The Idea of a Scientific Concept of Race". The distinction between a concept and conceptions of the concept is discussed in Hardimon (2003) following Tyler Burge and others. The same concept can be multiply articulated and these different articulations are referred to as particular conceptions of the concept. Here Hardimon proposed different concepts of race. Others may disagree as to his particular articulation of say, the ordinary, racialist, biological concepts and so would propose different conceptions for these concepts.

[13] Hardimon clarifies that "most common uses of the term "race" mobilize the concept of race in its racialized form", though he thinks that that "we use the ordinary concept when we use the term 'race'" (personal communication with the author, February 7th, 2009).

one's parents and one's parents' parents, etc. "A race is a lineage, a line" says Hardimon (445). The third criterion captures the association of the origins of these groups with different geographical -usually continental- regions, such as Africa, Eurasia, East Asia, America and the Pacific Islands.

Hardimon points out that the logical core does not include any criteria as to the "purity" of a founder population (449). Nor, says Hardimon, does the core require that each race be distinguished by properties particular to it and no other race (449). (I am assuming Hardimon here means no properties beyond those specified in HLC1-3 which are by definition part of the core properties that any concept of race will have and so specified on any populations correctly referred to as "races".) And because this is a concept applying to groups of humans, particular individuals need not satisfy all of the criteria to the same degree and some might satisfy these conditions in different degrees. So the ordinary concept of race in *its logical core* allows for cases of individuals that look indistinguishable but have a different ancestry or geographical origin, for individuals that look very different but have a common ancestry and belong to groups of the same geographic origin, and so on. The logical core is compatible too with essentialist conceptions of race[14].

Hardimon adds that the ordinary concept historically developed into the Racialist Concept of Race. Hardimon's *Racialist Concept of Race* may be specified as follows:

> The *racialist* development of the concept of race is a typological, essentialist concept of race. On top of the features considered in the ordinary concept's logical core (HLC1-HLC3), the racialist notion conceives races as human groups that differ according to features that I summarize as follows (452):

[14] See Hardimon (2003) for more detail on the importance of each of the three criteria in the logical core of the ordinary concept. How does one define the logical core of a concept? Should this stand in some logical relation to the rest of the properties of a concept? Or a logical relation to the context of use of a concept? If logical structure is important then does it matter whether the core features stand in different logical relations to each other? All these are important questions that Hardimon does not answer and nor will I. If logical structure internal to the core is of interest then perhaps only HLC1 should be placed at the core of the notion. HLC2 and HLC3 are logically secondary (relative to our concept of race) to HLC1 as they are often taken to explain HLC1 which is the manifest expression of race. Any structure internal to the core may be interesting to map, but it is not important for found science to proceed. So what I think about the concept does not matter here. I take Hardimon's definitions as a starting point.

+ (4) a *fixed* set of heritable physical, moral, intellectual and cultural characteristics

(5) an *essential* biological structure responsible for the strict correlation between physical features and inherited physical, moral, intellectual and cultural characteristics

(6) a *natural hierarchy* of inherited physical, moral, behavioral and intellectual characteristics

The racialist concept of race as specified by Hardimon is essentialist and typological; it presumes that characteristics under (4)-(5) are peculiar to each group and apply to each and every individual belonging to thus specified human groups.

Hardimon adds that the racialist notion has been debunked as lacking any epistemic basis and as functioning to legitimate unwarranted social-political hegemonies: It is ideological in what Raymond Geuss (1981) calls 'the pejorative sense'. Still, racialist conceptions of race remain popular and even if debunked have visible consequences which survive them.

Hardimon's *Populationist Concept of Race* can be specified next:

The *populationist* concept of race is a concept that Hardimon proposes is a scientific concept of race. It distinguishes population groups that share: (Hardimon's Populationist Concept, or HPC, criteria)

HPC (1) visually distinctive, genetically transmitted phenotypic characters of a relevant kind

HPC (2) a common biological lineage

HPC (3) an origin in a founding population that was initially geographically and reproductively isolated

Hardimon models this concept after Mayr's biological species concept and so they both share a generic notion of reproductive isolation. In the case of species the isolating mechanisms are "internal" to the species –typically consisting of biological traits, but in the case of races these mechanisms are posited as "external" and are typically represented by geographical boundaries. As with the other population-level concepts, differences described here are properties of the group and again not essential for particular group members; distinctive frequencies of characters

rather than the having or not of relevant features are what distinguishes populations. Hardimon adds that this is a notion intended as a "scientization" of the ordinary concept.

1.2.2 Hardimon on Socialrace

Finally, Hardimon defines an analytical concept of race, that is, a concept to be used by researchers wishing to analyze the consequences of racial classifications:

> *Socialrace* is a concept that distinguishes between human groups according to whether a group is *taken to be a race* in a specific society, where 'race' is conceived as *racialist* race.
>
> According to Hardimon, this concept is:
>
> (a) a *second-order* race concept: It requires the existence of a first-order race concept: the 'race' that human social groups are taken to be.
>
> (b) *analytic*: To understand the meaning of this concept you only need to look to what groups are taken to be racialist races in a society, and nothing more.
>
> (c) *social* and *structural*: It picks out kinds according to social context and situates these groups within institutional social structures,
>
> (d) *ideological* in the descriptive sense: It picks out 'ideology' in what Raymond Geuss (1981) calls the 'descriptive sense', i.e. discursive elements like our ideas and beliefs and so but also non-discursive elements like feelings.
>
> (e) *intensional*: It picks out its referent –its extension- according to what a social milieu or society takes (or intends –believes, etc.) race to be. If we distinguish between meaning and reference, in terms of intension and extension, this concept is engineered to keep track of the "meaning" of race, its intension, within shifting social milieus as opposed to what 'race' designates, or its extension. Socialrace tracks the intensions not the extensions of "race"-usage within particular contexts.

As Hardimon notes (ms1, 34) *socialrace* is a new name for a conception of race-as-a-social-group; a conception broadly used by race theorists studying race. Socialraces refer to entities that

belong to the realm of the social: they arise from the social organization of reality. Socialraces will vary depending on the social context in question. For example, the groups "white", "black", "Pacific Islander or Asian", "American Indian or Alaska Native" are U.S. socialraces but not Canadian socialraces[15].

The neologism is needed, says Hardimon, to distinguish this *social* notion from the ordinary, the racialist and the populationist notions that are all what Hardimon calls "biological in the basic sense" (ms1, 36). Though they are perhaps in possession of this concept race theorists have not been fully appreciative of the distinctiveness of socialrace Hardimon says and adds that there is a tendency to confuse the concept of socialrace with the idea that the racialist concept of race is socially constructed or with the ordinary concept of race —i.e. to read the concept of socialrace into the ordinary concept of race.

Hardimon calls the ordinary, the racialist and the populationist concepts of race biological in *the basic sense* because all three notions purport to refer to "things *biological"* –i.e. things belonging to the domain of living organisms (ms1, 14). So for example the concept of 'dragon' would also be biological in the basic sense even if there are no such living beings but the concept of a 'foe' might not (as a foe could be disembodied) and so would not be biological in the basic sense. But on what basis are we to deem a concept "biological in a basic sense"?

Note that concepts of race make distinctions between groups of humans and groups of humans are no obvious biological "things". What Hardimon means to say is that the ordinary, the racialist and the populationist concept make distinctions on the basis of criteria that point to biological features of members of these groups (ms1,14). These are the criteria that Hardimon locates in the core of the concept of race[16]. So here is a more accurate definition: a concept is biological in the basic sense if its logical core is defined in terms of biological criteria.

But then, in what sense is this biological, basic sense basic? If we take our instructions from the types of properties specified in these notions' logical core there would presumably be more ways that a race concept *could* be in a *basic* sense: that is, on top of more clearly biological

[15] Whether these groups are taken to be racialist races in the U.S. or not would be a matter of further empirical research though it is likely the answer is positive in the current racialist state.
[16] Personal communication with the author (February 7th, 2009).

features such as (1) visible physical features, the logical core specifies (2) ancestry and (3) geographical origin, as basic. So sure, perhaps the concept of race is biological in a basic sense, but perhaps also say ecological and historical, or even sociological in a *basic* sense too; all depending on how we could understand 'ancestral lineage' and 'geographical origins' in a particular context. In this case we might end up positing a group that is not only biological –but also, and contrastingly in Hardimon's frame– *social* in a basic sense.

Given that the notion is so characterized on the basis of the properties specified in its logical core, the possibility of the concept of race being more than *biological* in a basic sense is a live one. By emphasizing the fact that race as defined here are biological in a basic sense Hardimon is "hammering home" a message that he quite rightly says is often lost to those studying race. But this is not the only correct message one could communicate[17]. A choice of what is taken to be 'basic' seems to depend to some extent on what message is interesting to hammer home.

In any case, let us sidestep the issue momentarily. Hardimon defines 'socialrace' as a concept that is not biological in the basic sense. It captures kinds that are of a social, analytical, intensional (etc.) type.

1.2.3 Why Less Can Be More

The specification and distinction between four race concepts offered by Michael Hardimon is certainly of use in clarifying a discourse on "race" that is often fraught with misunderstandings and conflations of different notions. Would these concepts in fact describe how users of "race" in the current, and recent, U.S. context conceive of race? Haslanger (2005) doesn't think so. She takes the multi-use of "race" to render this a moot point.

But what is described by Hardimon is not necessarily individual usage of the term "race" but rather general concepts of race –I take it that his aim is to provide some conceptual

[17] See Kitcher (1999) for how races might be taken to be both biological and social. Or Hacking (2006) on biosociality. Neither of these accounts deal with the concept of race –but rather with the classes thus classified (in the case of Kitcher) and with how the process –or rather the fact– of classification matters (in the case of Hacking).

scaffolding that individuals' concepts might build on or develop from. His is an approach from the inside out, so to speak. From a core to criteria, to a concept.

This dissertation follows an opposite direction. It starts with surface: "race", the word, and its usage in text. The analysis doesn't hit on any cores: that is I cannot prove that the concepts associated with different usage of the term "race" map onto the concepts that Hardimon specifies. Though were we to assume that such conceptual cores –or at least conceptual structures– were in place then found science points to certain conclusions. Namely, that race notions can get transfigured and come to function as notions with totally different cores: scientific ones[18].

My approach is different because my aims are different. The aim of this dissertation is not to specify what concepts of race or race concepts people could *have* in the current U.S. context but rather to describe how concepts of race *vary* and how they come to vary specifically across ordinary and scientific contexts; and to describe what these concepts are by virtue of being so varied and variable. To study variation one must use some fixed points. Hardimon's formulations serve as initial guides because they could be so fixed, as concepts representative of the variation of "race" usage. But they point to a more general pattern. And the concepts specified by Hardimon end up being too detailed to be useful for picking out the pattern found science points to[19].

This is why I define vague race concepts. To use a physics analogy again, one may derive equations for a fluid's flow by reference to a particular particle's motion in that fluid, but WHICH one this particle is is not important; not specifically; only insofar as it is a part of this fluid does it help to look at *it* and describe *its* motion. Similarly for WHICH criteria specify a race concept: criteria specifying a race concept are important in the context of this project only insofar as they relate to the variability of the usage of race concepts across ordinary and

[18] There are many metaphors I used while thinking about these transfigurations. My initial one was borrowed from physics where I received my undergraduate training. This was the idea that an electron can behave –and for all purposes BE –a wave or a particle when we use different methods to detect it –the analogy being with 'race' being understood multiply across different disciplinary contexts by virtue of this concept entering different contexts of use with ontologies and instruments already in place for such type of thinking. But I think that at its basic sense the analogy I provide in section 1.4 by the possibility of using a pen-cap as a hair-clip captures this same type of variability. And I hope it is easier to entertain than that of detecting electrons via different methods!

[19] See section 2.3.1.1 for a general argument for why "superficial" properties may be useful

bioscientific biomedical ones. And they don't –not directly. What matters is not WHICH criteria these are, but rather, that they are taken to be specifying 'race', as understood within particular, social or bioscientific, contexts.

My choice "dyes", so to speak, to study the use of "race"-tagged concepts (i.e. understandings of the term "race") in a biomedical context are the second order race concepts: 'sociorace' and 'biorace' which are defined in the next two sections[20]. Sociorace and biorace are engineered to capture any changes to a concept associated with a background change, by virtue of being defined *relative* to a (possibly mutable but identifiably) social or scientific context. They are perhaps less clear but clear in the right ways for tracking what features of "race"-usage this dissertation proposes it is interesting to describe.

1.3 Biorace and Sociorace Race Concepts

In his essay on "The Ordinary Concept of Race" Hardimon asks: "Why is a philosophical investigation of the ordinary concept of race called for?" he motivates his analysis as follows (Hardimon 2003, 438):

> Researchers in other disciplines quite properly approach race as a concept within their own fields. Biologists take the concept to be biological. Social theorists take it to be social. Philosophy on the other hand, provides the resources to explore the concept without prejudging its disciplinary status (Hardimon 2003, 439).

It is with this seemingly passing observation that I preoccupy myself in this dissertation. Hardimon puts his finger on something that sounds pretty commonsensical: there seem to be discipline-specific understandings and correlative uses of the term 'race'. But why? And more importantly, (as Nancy Cartwright asked me) why these and not others?

I argue that there is a process whereby ordinary concepts could come to be embedded in a scientific context. I call this process a process of "finding" and "founding" ordinary concepts in

[20] As mentioned already the possibility of using these notions was made visible to me by Hardimon's concept of socialrace; hence the similarity in the structure of socialrace with sociorace and biorace concepts.

science (Section 1.4 in this chapter), and I call its results "found science" by analogy to "found art" (Chapter 3, Section 3.1).

But first, let me draw a distinction which when applied will help notice the pattern I am referring to. This pattern is observed by picking out two second-order race concepts that I call "sociorace" and "biorace" concepts.

1.3.1 Sociorace Concepts

I define **sociorace** concepts as follows:

> *Sociorace* is a concept that distinguishes between human groups on the basis that they are taken to be 'races' in a particular social context[21].

Sociorace is a second-order race concept; it distinguishes between human groups in terms of a first-order race concept –the 'race' groups are taken to be; a concept which could be the ordinary one, or the racialist one, or some other race notion.

Sociorace is modeled after Hardimon's definition for socialrace, with an open slot for where the first order 'racialist race' notion went; and a specification of the type of intensional context we're in. It is

(a) a *second-order* race concept. It requires the existence of a first-order race concept: the 'race' that human social groups are taken to be.

(b) *analytic*: To understand the meaning of this concept you only need to look to what groups are taken to be 'races' in a society, and nothing more.

(c) *social* and *structural*: It picks out kinds according to social context and situates these groups within institutional and social structures,

(d) *ideological* in the descriptive sense: It picks out 'ideology' in what Raymond Geuss (1981) calls the 'descriptive sense', i.e. discursive elements like our ideas and beliefs and also non-discursive elements like feelings.

---

[21] **Socioraces** are, in the language I introduce in Section 1.4 of this Chapter, **founded** races, what we'd use as a first-order race notion as founded in social contexts of space and interests. The same human group founded in different social science contexts may correspond to different socioraces.

(e) *intensional*: It picks out its referent –its extension- according to what a social milieu or society takes (or intends –believes, etc.) race to be. If we distinguish between meaning and reference, in terms of intension and extension, this concept is engineered to keep track of the "meaning" of race, its intension, within shifting social milieus as opposed to what "race" designates, or its extension. Sociorace tracks the intensions not the extensions of "race"-usage within particular social contexts.

Unlike Hardimon's definition for *socialrace*, this definition leaves a first order race concept unspecified. *Sociorace* describes a set of concepts with the properties of socialrace and to which socialrace would belong but it is defined with an unspecified first order notion at its core, to leave room for other understandings of 'race' in social contexts other than a U.S. one to also be thus understood, as socioraces.

When I use the term sociorace I will use it as a noun, so I will say that 'socialrace is a sociorace', but also as an adjective: By a 'sociorace concept of race' I understand a 'concept of race as a social kind'. Socioraces are by definition groups *taken to be* 'races' within particular social contexts.

1.3.1.1 Sociorace Concepts Need not Pick Out Races

If one accepts a. that the ordinary concept of race has a logical core and b. that the ingredients of the logical core of race concepts are those proposed by Hardimon, then some socioraces can be defined operationally as concepts that attempt to articulate conditions in the *logical core* of the ordinary race concept, that is,

HLC (1) visible physical features of the relevant kind,

HLC (2) common ancestry and

HLC (3) distinctive geographic origin,

into socially meaningful terms. (Though other non-"essential" features of a race concept could also be thus articulated.)

So, for instance a sociorace notion of race may translate visible physical features into a dress code, or translate common ancestry into shared family, or translate geographic origin into that of a country of origin[22]. (Such an operative notion could be more akin to what many understand by "ethnicity"). Note here, that I am not requiring that all the HLC criteria be re-articulated at the same time by each sociorace notion. This could be a divergence from Hardimon's understanding of a logical core as such; as core; as necessary; though such modifications would also leave open the question of what sociorace notions are notions of, properly speaking and whether or not they are concepts of race.

Take for example the work that the U.S. Office of Minority Health (OMH) has to do. OMH was set up in 1986 "to improve and protect the health of racial and ethnic minority populations through the development of health policies and programs that will eliminate health disparities"[23]. OMH is housed within the U.S. Department of Health and Human Services (HHS) and its duties are various including running health campaigns like "A Healthy Baby Begins with You", or "Celebra La Vida Con Salud" (Celebrate a Healthy Life), or "Take a Loved One to The Doctor Day"[24].

These campaigns involve communicating knowledge about health issues to the communities that are taken to be at higher risk of suffering from them. For example "A Healthy Baby Begins with You" is specially focused to reach African-American communities whereas "Celebra La Vida Con Salud" (Celebrate a Healthy Life) is targeted at informing Latino communities. We read on the relevant website that

> *Celebra La Vida Con Salud* promotes prevention, helps link Latinos to health services, and celebrates health by drawing upon family and community strengths to make good health practices a part of daily living. The campaign is a series of educational events and media outreach activities including a traveling health fair

[22]Hardimon notes that such translations need not modify the core but rather only add these supplemental features to it. He is right to point this out. However, found science proposes that systematic use of these "dressed-up" notions is expediently only going to stay on the surface. That is, there might be a core to the notion that Hardimon proposes, but for all we know and care, what matters is the found use of the concept, and this comes with the appropriate "dress". In this case and in the language I introduce in Section 1.4 of this Chapter, a found dress code would be a **founded** "visible physical feature of the relevant kind", found family a **founded** "ancestral group" and found national origin a **founded** "geographic origin", each founded in social contexts.

[23] http://www.omhrc.gov/templates/browse.aspx?lvl=1&lvlID=7 [Last accessed 30/9/08]

[24] Ibid.

> touring 10 cities, health segments on the *Prevenir es Salud* and local radio shows, and public service announcements featuring health messages on issues that impact the Latino community. To date, thousands of people have participated in the health fairs and received free health screenings and information. Stay tuned for a fair in a city near you[25].

Such health programs may be communicating knowledge about biological features or even genetic risks that concern these populations. However the concepts of 'race' and 'ethnicity' agents are utilizing to carry through the goals of these practices are not ones of race as a biological kind but rather of race as a social kind.

For example, what aesthetic motif on a poster or leaflet might attract these patients? What language should be used to communicate with patients and their families or carers (and practice "*cultural* competence", in the terms of biomedical ethics)? What community ties will be useful in disseminating what knowledge needs to get through to individuals concerned? What institutional or community structures may help sustain engagement with these groups? All these questions are the types of questions that the implementers and designers of health policy as well as medical practitioners need to ask and answer to bring through the medical targets of their work. And all these rely on an understanding of race/ethnicity categories as social (cultural, political, etc.) kinds. (And these are only some of the questions people would in fact have to ask to get such initiatives off the ground and working.) So, what following Haslanger we would call an operative concept of race is in this case a concept that understands race as a social kind. Which is what I want to tag by the general category of sociorace.

These questions can be thought to rely on concepts that articulate criteria in the core of race concepts in socially meaningful terms. The notion of distinctive physical features may be extended to include distinctive features of appearance or presence which may include colors of dress, smells or sounds; the notion of a shared ancestry can be understood as that of a shared community with its own norms and practices; and that of a distinctive geographic origin as that of a distinctive social, political origin, a distinctive history from that of the U.S. context. Note that not all of the criteria in the core of the concept would have to be articulated in each case. In some

[25] Ibid.

instances, for some purposes, thinking about community ties as associated to 'race' will be of interest in isolation from thinking about language or aesthetics or economic background or something else. And in some cases what criteria are rearticulated criteria need not even correspond to core criteria of an ordinary race concept.

Sociorace concepts would also be the types of concepts used to explore what and how social/ environmental factors are causally efficacious in shaping health outcomes across human groups. (An argument for this claim is offered later in the dissertation, in Chapter 2 Section 2.1.4) The medical and philosophical discourse acknowledges the existence of social correlates of race as confounds rather than causes. But I think there is more to be said for understanding race itself as a social kind in such endeavors. The first-order race concept that describes what 'races' groups are taken to be could be racialist race, ordinary race, biological race, or some other notion[26]. Sociorace, like socialrace, need not *itself* be a racialist or essentialist concept as it is one degree removed from what first-order race notion might be used to define it. But it would be a useful type of notion to use to track such possible effects of racism (from this [more] safe distance). Of course, care needs to be taken when defining sociorace notions for use for social scientific use: not all sociorace notions will be scientifically significant.

1.3.1.2 Some Sociorace Race Concepts: Hardimon's 'Socialrace' and Haslanger's 'Race'

A sociorace notion that bears striking similarities to Hardimon's notion of socialrace is the concept Sally Haslanger proposes for 'race'. Sally Haslanger defends a social constructionist view of race and gender; which is not to say, according to Haslanger, that the categories are unreal nor to merely state their historical origins but to claim that they track social structure and how social privileges are distributed better than biological structure (though in cases they may be tracking –and tracking well- only that). Haslanger defines race as the social meaning of color (by analogy to her definition of gender as the social meaning of sex), where 'color' (like 'sex') refers

[26] The first order concept could be a sociorace or biorace notion itself, though at (the core of) its core would have to be a biorace race notion, I think.

to the physical features on the basis of which individuals' social positions are "marked". In her words:

> I will use the term '*color'* to refer to the (contextually variable) physical 'markers' of race, just as I use the term 'sex' to refer to the (contextually variable) physical 'markers' of gender. I mean to include in 'color' more than just skin tone: common markers also include eye, nose, and lip shape, hair texture, physique, etc. And in principle I want to allow that virtually any cluster of physical traits that are assumed to be inherited from those who occupy a specific geographical region or regions can count as 'color'. (Although the term 'people of color' is used to refer to non-Whites, I want to allow that the markers of 'Whiteness' count as 'color'.) (Haslanger 2004, 6)

She defines a group G as racialized relative to a particular context, as follows:

> A group G is *racialized* relative to context C iff members of G are (all and only) those:
> a. who are observed or imagined to have certain bodily features presumed in C to be evidence of ancestral links to a certain geographical region (or regions);
> b. whose having (or being imagined to have) these features marks them within the context of the background ideology in C as appropriately occupying certain kinds of social position that are in fact either subordinate or privileged (and so motivates and justifies their occupying such a position); and
> c. whose satisfying (1) and (2) plays (or would play) a role in their systematic subordination or privilege in C, that is, who are along some dimension systematically subordinated or privileged when in C, and satisfying (1) and (2) plays (or would play) a role in that dimension of subordination or privilege (Haslanger 2005b, 271).

I should caution here that Haslanger sees an important difference between Hardimon's project and her own. This definition is for Haslanger a tool for opening up people's eyes to what race "really" is, in most social contexts, today. It is a normative, analytic notion. But this doesn't mean there is no overlap in the work of Hardimon and Haslanger; nor that Hardimon's account comes without what Haslanger would call "ameliorative" ends.

Haslanger defines race as the "social meaning" of a notion (color) that is biological in the basic sense. Haslanger's definition is structured like that of socialrace. She posits a first-order notion (color) given meaning by a social context, which defines thus a second-order notion (race), just as Hardimon posits a first-order notion (racialist race) given meaning by a social context thus defining a second-order notion (socialrace).

Further, Haslanger defines race as a social status that can track trajectories of power. How? By following "markings" left by the first-order (color) notion on the bodies of social subjects, which are just like the ones racialist race would leave. What marks would these be? a.

Physical features of the relevant kind for Hardimon or "eye, nose, and lip shape, hair texture, physique, etc." for Haslanger, and b. posited moral or other differences in capacity for Hardimon (that are ideological in the pejorative sense) or physical marks as interpreted and deciding who is "appropriately occupying certain kinds of social position" for Haslanger.

It is likely therefore that one could build a correspondence between Haslanger's 'race' and Hardimon's 'socialrace'. How? Build a correspondence between Haslanger's notion of color and Hardimon's racialist race notion and then follow that up with a correspondence between Haslanger's notion of race and Hardimon's notion of socialrace. I am specifying racialist race instead of ordinary race as the first order race notion we might want to use in the sociorace (socialrace) definition because this can keep track of the content attributed to race by racist ideologies. Both Hardimon's socialrace and Haslanger's race take into consideration the use of racial classifications to legitimate unwarranted oppression (excess *Herrschaft* in the terms of Raymond Geuss[27]) and both are proposed by their authors as analytical tools in order to track such distributions of hegemony in a society. Haslanger wants to reserve the term 'race' for the "critical analytical" or "ameliorative" job that she takes her project to do where Hardimon uses a neologism to name his notion, which works much better to keep our terms straight –even if it doesn't reverberate with the sentiments the ordinary word arouses.

I should note that I am making building this correspondence sound slightly more straightforward than it is. The difficulty arises with a difference in how Hardimon and Haslanger understand concept-building. Hardimon speaks of logical cores and lists criteria external to a notion's core while Haslanger defines notions operationally and contextually. So, her notion of 'color' is much slimmer than Hardimon's 'racialist race' and the racialist overtones that color gets is understood as a social background effect not explicitly listed as a criterion in a specification of the concept's contents.

If such a correspondence is not built, Hardimon might have to explain why Haslanger's notion is a race notion; why it satisfies the criteria in the logical core of the ordinary race concept

[27] Geuss (1981)

that Hardimon specifies as holding for any race concept. Hers is a concept that doesn't make direct reference to HLC1-HLC3; it only makes reference to 'color' which approximates HLC1, and posits that HLC1 is understood as ("taken to be") linked to HLC2, and HLC3, within a particular social context, thus seemingly moving these core criteria further away from the logical core where Hardimon puts them. Similarly, Hardimon would have to account for why other sociorace notions, like the operative notions of 'race' in the public health examples mentioned, are concepts of race[28].

But this is not the only option. Hardimon could say that these are not properly speaking concepts of race –they don't satisfy the HLC criteria. He then would need to explain why and how the label 'race' may come to be used to tag these not-'really'-race-concepts. Or Hardimon may need to re-examine the ingredients of the concept's logical core. Or he can do both!

This is what I try to do by articulating a process for finding and founding an ordinary notion in a particular context.

1.3.2 Biorace Concepts

The other second-order race notion I define is **biorace.** I define **biorace** as follows:

Biorace is a concept that distinguishes between human groups on the basis that they are taken to be 'races' in a particular bioscience context[29].

Biorace is a second-order race concept: it distinguishes between groups in terms of a first-order concept of race. And the definition for biorace is modeled after the definition for sociorace so bioraces share some of the features specified for sociorace, again with an open slot for where the first order 'race' notion goes, and a specification of the type of intensional context we're in.

Biorace is

[28] Hardimon thinks socialrace just is what Haslanger's notion of 'race' is (personal communication, February 7th, 2009).

[29] **Bioraces** are, in the language I shall introduce in Section 1.4 of this Chapter, **founded** race concepts, what we'd use as a first-order race notion as founded in bioscientific contexts of space and interests. The same human group found in different bioscience contexts may correspond to different bioraces.

(a) a *second-order* race concept. It requires the existence of a first-order race concept: the 'race' that human social groups are taken to be.

(b) *analytic*: To understand the meaning of this concept you only need to look to what groups are taken to be 'races' in a bioscience context, and nothing more.

(c) *biological* and *theoretical*: It picks out kinds according to properties taken to be biological and situates these groups within bioscientific theoretical structures,

Biorace is a notion that is also:

(d) *ideological* in the descriptive sense: It picks out 'ideology' in what Raymond Geuss (1981) calls the 'descriptive sense', i.e. discursive elements like ideas and beliefs. Though these are beliefs regarding bioscientifically meaningful terms.

(e) *intensional*: It picks out its referent –its extension- according to what a social, bioscience milieu takes (or intends –believes, etc.) race to be. If we distinguish between meaning and reference, in terms of intension and extension, this concept is engineered to keep track of the "meaning" of "race", its intension, within shifting bioscientific theories and practices as opposed to what "race" designates, or its extension. Biorace tracks the intensions not the extensions of "race"-usage within particular bioscience contexts.

Again this definition leaves a first order race concept unspecified. It is vaguely defined to leave room for different understandings of 'race' in different bioscience contexts to also be thus understood as bioraces. When I use the term "biorace" I use it as a noun, so I will say that "Hardimon's 'populationist race' is a biorace", but also as an adjective: By a 'biorace concept of race' I understand a 'concept of race as a biological kind'.

1.3.2.1 Biorace Concepts Need not Pick Out Races

If one accepts a. that the ordinary concept of race has a logical core and b. that the ingredients of the logical core of race concepts are those proposed by Hardimon, then bioraces can be defined operationally as concepts that attempt to articulate conditions in the *logical core* of the ordinary race concept, that is,

HLC (1) visible physical features of the relevant kind,

HLC (2) common ancestry and

HLC (3) distinctive geographic origin,

into biologically meaningful terms. Though criteria outside the core could also be thus articulated.

Many biorace race notions, even in the case of human races, are biologically respectable though none is agreed on as biologically significant. Biorace notions need not be typological or essentialist and they can be useful for tracking biologically interesting difference and asking whether there is medically significant genetic variation which is stratified according to race/ethnicity.

Examples of biorace concepts would include Hardimon's populationist concept of race which retranslated the core notions HLC(1)-(3) into these corresponding ones of (Hardimon's Populationist Concept, or HPC, criteria)

HPC (1) visually distinctive, genetically transmitted phenotypic characters of a relevant kind

HPC (2) a common biological lineage

HPC (3) an origin in a founding population that was initially geographically and reproductively isolated

But consider also Andreasen's *cladistic* race concept (Andreasen 2000) **or** the *ecotypic* race concept of Pigliucci & Kaplan (Pigliucci & Kaplan 2003). Andreasen defines race on the basis of *phylogenetic* features**.** On his account races must share features HPC (2) and HPC (3) [clades are branches of a phylogenetic tree], but visible differences of the relevant kind are *left out* of the defining features (presumably, but not necessarily caused by HPC (2) and HPC (3)). On Andreasen's account the group *Asians* would not form a monophyletic group, so would not be a cladistic race (-nor would Europeans).

Pigliucci and Kaplan on the other hand emphasize *phenotypic* features**.** On their account races are genetic adaptations to environmental conditions. This definition emphasizes feature

HPC (1) and HPC (3), *leaving out shared genetic lineage* as a feature members of the same race must share, as the same ecotype may be noticed in populations of different origins.

Both Andreasen (2000) and Pigliucci and Kaplan (2003) propose biologically respectable accounts of 'racial' difference but the races they specify don't include all three features Hardimon locates in the core of the concept. Andreasen picks out features HLC (2) and HLC (3) in the logical core as basic; whereas Pigliucci and Kaplan focus on HLC (1) and HLC (3). Can they legitimately do this, if we accept Hardimon's account of the ordinary concept? How and why would they?

As before, this is not fatal to Hardimon's account of the concept of race. One could dismiss these concepts right off: "I don't know what these are concepts of, but they are not concepts of (ordinary) 'race'". Though again Hardimon would have to account for why we then call them 'race'. On the other hand one may also question Hardimon's choice of criteria and/or their location; whether these should be put in the logical core of a concept of race.

I choose to stick with Hardimon's criteria and give an account of why and how, though these are not ordinary race concepts, they are still tagged by the word 'race'. I take the possibility of thus rearticulating an available notion to challenge Hardimon's claim that this logical core is really a hard one. But this is not a discussion of present interest.

This apparent problem for Hardimon's account of the ordinary race concept helps highlight the need for an alternative understanding of concepts or what conceptual articulations amount to in practice: Though I accept Hardimon's hypothesis that there are some (3) principles in the core of any race notion, I conclude that these can be transfigured logically, "rearticulated", or "translated", in a way that may ultimately change a notion.

I explain how I understand this to happen in the following section. Its results I call "founded" race notions; though founded notions can also get to be named depending on the target domain in which are founded, as "found _____ (risk factors, genetic populations, etc.)".

1.4 Found Concepts and Founded Concepts

I distinguish between a concept that is "found" in a particular context of use and a concept that is "founded" in that context. These two terms carry a special meaning. They correspond to two parts of a process for embedding a given concept in a particular context of use: 1. finding a notion as available, but ordinary, in a scientific context and 2. founding it there, as scientific.

An ordinary notion should be one we can find in an ordinary context of use. By which I mean a context of ordinary discourse. What is necessary for founding a concept in a scientific context is that it be found in that scientific context, as ordinary; this would be a first step towards founding the notion in a scientific context[30].

Finding and founding must usually happen recursively, at different levels within a context, before a concept becomes embedded in a context. And it doesn't happen randomly: it depends on the particular specifications of the concept, the interests and concepts of users in the target context and the tools available for the embedding –tools that I call "founding tools".

Corresponding to these two phases of the process I call the concept 1. "found" in the target context and 2. "founded" there.

Chapter 3 of this dissertation examines specific examples of finding and founding ordinary race concepts in the context of biomedical genetics. But this is an instance of a process that I think is general and happens all the time, featuring entities that are much more mundane than floaty concepts and target contexts more banal than scientific ones. The following analogy shows the simplicity and ubiquity of finding and founding processes.

1.4.1 Finding and Founding

I will first describe what it is to "find" and "found" something in a particular context of use by using a mundane example. This is the example of a pen-cap found and founded as a hair-clip.

[30] Saying that the ordinary notion must be found in a scientific context seems to say more than it is. Given that agents in scientific contexts also inhabit ordinary contexts and that our everyday practices are intertwined intersections of ordinary contexts and scientific ones where these "findings" would occur are much more common than it would first appear

1.4.1.1 Take A Pen-Cap

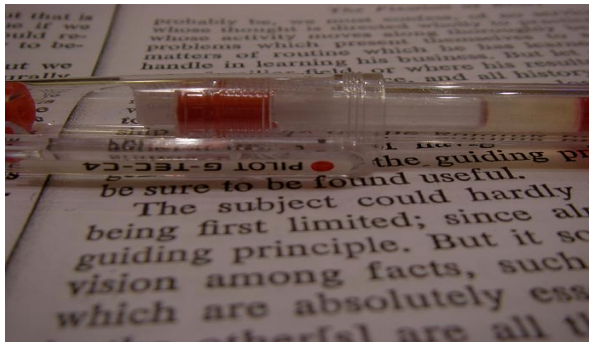


Figure 1: A pen with its pen-cap on.

This pen-cap has a particular design. Certain features of a pen-cap's design serve particular functions which users and producers of caps of pens have deemed useful.

These pen-cap features can be organized as features for

i. *Preserving:*

a1. the function of the pen by

a1.1 making the pen-cap hollow and sturdy to protect the tip of the pen, a1.2 making the pen-cap's rim fit the pen tightly, or

a2. the shape of the pen by

a2.1 making the pen-cap of the same shape as the pen, etc.

and features for

ii. *Exporting:*

b1. the pen by

b1.1 giving an arm to the pen-cap so we can stick the pen in places, or b1.2 making the cap look pretty so we sell more of them

b2. the pen-cap by

b2.1 making the pen-cap a standard shape so it fits other pens, etc.

A pen is also usually responsively manufactured to sustain a relationship between itself and its cap: a pen's shape is usually back to front symmetrical so that a pen-cap may be stuck at the back of the pen when it's not covering a pen's tip, to avoid losing the pen-cap.

These are features that the pen-cap pictured in figure 1 has and you will agree that it has these features and not others, in line with particular interests of users and producers of "pen-caps". What one usually does with the cap of a pen is enabled by such features though not necessitated by them.

Which brings me to the possibility that finding and founding hinges on.

Were you to come across the object depicted (figure 1) and given that you are familiar with the category of "pen-caps", what you would most likely do with this object is use it in one of the ways specified: you would probably look for a "pen" in the pen-cap's vicinity and stick it in place, or you would possibly put the cap in a pencil case, or leave it there, on your desk, or some such.

But this is not all that one can do with a pen-cap.

Depicted above was the cap of my pen.

It looks like any other pen-cap. But it so happens that it is mine. And it so happens that I have a fringe; and that when I write and look down to the paper, this fringe gets in my eyes. It also happens that when I write I often have a pen and not a hair-clip on me.

So here's a habit I have developed. I use some of the caps of my pens to hold my fringe away from my face. How? If the arm of a pen-cap makes contact with the body of the pen-cap and there is enough space between the two for my fringe to fit, then that type of pen-cap can be used as a hair-clip, to hold my fringe up and out of my face.

Figure 2: Things labeled “pen-cap” and things labeled “hair-clip” have certain capacities in common: holding up S’s hair.

This is no trick. It happens because of a real capacity that this type of pen-cap has; a capacity that was not foreseen by its designers, but one that is of real use to a user: me.

Perhaps I am annoying you, talking about my self, and my hair, etc.

-What does this have to do with race concepts and science?

Now, bear in mind that I do this now, all the time, with this particular sort of pen-cap. So even in this simpler case the question can arise: What is this?

-Is it a pen-cap?

or

-Is it a hair-clip?

[-It’s super-cap!]

Answer: It is not a pen-cap and it is not a hair-clip.

It is a “founded” pen-cap.

And it is a “found” hair-clip.

How? It is a pen-cap found and founded as a hair-clip.

A general description for how x can be found and founded as a y is given in chapter 3. As a proto-definition we might say that x is a *found y* when x is used as if it were a y in a practice of y-use; and x is *founded as a y,* when x is recognizable as a y in a practice of y-use, where the practice of y-use happens repeatedly and tacitly –at no cost of translation.

1.4.1.2 A Founded Concept

What is important in this story is the process of finding and founding. But turn to 'concepts'. Two ways were used to specify the object of the example: "it is a 'found hair-clip'" and "it is a 'founded pen-cap'". These are results of a process which can operate on concepts.

What happens to my concept of 'pen-cap' as I keep finding use for many caps of my pens as hair-clips?

Well, I certainly haven't *lost* the concept of a pen-cap. I can still respond to a request like "Give me a pen-cap" quite successfully giving you the expected object. I can use the word in conversation, etc.

However I see pen-caps now, and I no longer *just* see pen-caps. In fact, what I do is that, often, when in need of a *hair-clip* I go into my bag or pencil case, and I *look for* a pen-cap. I will unscrew the cap of one of my pens and clip my hair. Whether or not I am writing at the time[31].

The founded concept I have of a pen-cap, in that context works for me, operates and picks out for me, what using the concept 'hair-clip' would: it is a found hair-clip concept or a founded pen-cap concept.

So, once I am in the context of use of the concept 'hair-clip' (like when needing one), and given the practice I have been describing as one of mine, what is to others a pen-cap is now understood and used by me as a 'hair-clip'; though it may still be tagged for all intents and purposes by the ordinary concept 'pen-cap' too.

I say what "*was*" a pen-cap because, though nothing might actually happen to the physical structure of this pen-cap *initially* you can imagine that if I keep using the pen-cap as a hair-clip and the hollow part where the pen tip should go gets clogged up with dirt, and the connection with the pen it came from gets lost because it is not *functioning* as a pen-cap, etc, that the pen-cap might actually come to only *be* usable as a hair-clip. And I might forget it ever was a pen-cap. Getting stuck in my own particular practice of just using this as a hair-clip could result in me losing the notion of *this* as pen-cap.

[31] This is a risky business as I end up with loose pens damaging my bag and also risk losing the pen-caps.

It is part in fear and part in joy with this possible disconnect that I propose to track it. I call this way of specifying what this object is, i.e. it is a ‘found hair-clip’ and a ‘founded pen-cap’, a concept; and as it is the result of the aforementioned process that I would call it a founded concept.

1.4.1.3 Take A Race Concept

How I got to use a pen-cap as a hair-clip is similar to how one gets to use a concept of race as a scientific variable. A race concept can be a found scientific variable and a race concept founded in a scientific context will not properly speaking be an ordinary race concept –if a race concept at all. Specific examples will be discussed in Chapter 3 but here is I hope enough to convince you that these two cases are comparable.

What typifies a found object is its availability and the possibility and convenience afforded by using it to achieve a purpose apparently unrelated to it. Let me remind you of the setup. In the pen-cap example am writing so performing an action which requires I that I be already using a pen. While writing my fringe falls in my face, and I need a hair-clip. What is responsible for me using a pen is coincidentally responsible for me needing a hair-clip: writing. This makes it possible that there is too a common solution –using part of the pen, not used for writing, as a hair-clip: finding and founding its pen-cap as a hair-clip.

Now, move to a biomedical setup. (This example will be terribly simplifying so please see chapter 3 section 3 for more detail regarding this type of thinking).

I am now a doctor, an epidemiologist working in the context of clinic C, in the social context of society S. As a matter of contingencies having to do with my upbringing and the history of the social context I inhabit, I am a competent user of an ordinary concept of race –say the one specified by Hardimon (2003). At the same time and as a matter of my biomedical, epidemiological training I am also a competent user of notions such as ‘incidence rate’, ‘prevalence’, ‘significant association’, ‘regression model’, ‘interaction term’, ‘subgroup analysis’, ‘risk factor’, etc.

Now, say I am tasked with studying disease D. One way I am trained to do this as an epidemiologist is by measuring the incidence rates of a disease D across different groupings of my patients and coming up with some prediction about what features that a patient may have – and by which I sorted them into groups– are properly speaking "risk factors" for developing disease D.

How would I normally do this? Roughly speaking I would divide my patients into groups in different combinations and look using statistical tests to see which set of which of these divisions will be "significantly associated" with developing D. "Subgroup analysis" is how I look to see which one of these features is important for responding to developing disease D. I call each of these features a "risk factor" for D.

What I notice, in particular setup, is that besides features that I have been trained as an epidemiologist to measure such as a patient's blood-pressure, or their Ankle-Brachial Index (ABI) or their Body Mass Index (BMI) or such, it happens that as a matter of social fact, my patients also come to me "sorted" –self-sorted even– by an ordinary concept of race[32]. So, a concept that allows me to sort my fellow citizens (and myself) into 'races' also allows me to sort my patients into groups. So, I also use that grouping in the set of groupings I examine. I divide my patients into groups by making use of a concept of 'race/ethnicity'.

And here is the surprising and useful outcome: What excess risk I hoped to "capture" by the different variables I consider I can capture using this grouping: "race"! Sorting my patients by 'race/ethnicity' picks up significant variations in my patients' risk for developing D. After multiple analyses, analyses where I look for other factors that I suspect might be associated with 'race/ethnicity' and responsible for a higher incidence of D across these groups, and looking for interactions between these –say I do this using a "multivariable logistic regression model" and a "model with interaction terms for race and each of the other risk variables" – I notice that D is

[32] As Appendix 3 argues a race concept found in a U.S. epidemiological setting would not properly speaking be ordinary; it would already be founded in demography. So these –as Michael Hardimon pointed out to me, would be racial or ethnic categories or labels, not ordinary concepts of race. I am not clarifying this here for simplicity as we can think of race concepts used in demography as still "non-scientific" or "common" relative to an epidemiological setting.

consistently more common or “prevalent” in patients who self-identify as of one of the race/ethnicities, say R.

I conclude that: race R is a consistent and independent risk factor for D at a magnitude similar to that of other established risk factors for D. –The reason why this is a significant association is unclear. But there is a direction suggested for me to look for causes of D: look among conditions typical to a group of people of race R and not of a group of people of race not-R, in C, in S.

So what is this “race”?

-Is it a concept of race?

-Is it a risk-factor?

I say it’s neither.

It is a “founded” race concept.

And it is a “found” ‘risk-factor’ for disease D, in the scientific context C, of society S.

1.4.1.4 The Risks of Using Founded Concepts

So far this was a happy story; almost. Truth is, when using my pen-cap as a hair-clip I made a choice. I could have decided to keep using my pen-cap as a pen-cap and stick the whole thing, pen and pen-cap to my hair. But this would be a bad solution to my hair problem. The whole thing would just be dragged down by the weight of the pen, and presto, the fringe would be in my face again. Plus using the pen-cap as a pen-cap and a hair-clip would negate the reason I needed a hair-clip in the first place: if I was not writing with the pen (rather had it stuck in my hair) then I wouldn’t really need a hair-clip to begin with.

A tricky situation indeed!

But one interestingly similar to how a concept of race might come to be specifically usable as a scientific concept in a scientific context. That is, we might often be thinking we are being or should be proper and respectful in our usage of the Concept of Race in our use of it in science; but we’re often not and we really shouldn’t be.

As defined by Michael Hardimon, there are three main principles in the logical core of the concept: HLC1-3. The criteria HLC1-3 in the core of the concept of race are list-able and so I think move-able to different symbolic contexts. –Even if you didn't really have a clue about why a pen-cap has an arm (was it born with one or was one constructed?) you could *still* be using that pen-cap as a hair-clip. This is an important fact about the matter. Similarly, knowing how or why what is on a list of criteria in a concept's logical core or not got to be on that list is not necessary for use to be made of it. It is possibly useful for helping to locate the right concept for one's purposes to be aware of candidate concepts' genealogy; plus it helps delineate the possible uses of a concept and helps decide what counts as a good or bad use of a concept to be aware of its history. So, it may help us as analysts to make a list of criteria in a concept's logical core to understand what is going on where and what can legitimately be going on where; though even for the sake of analysis sometimes a coarser-grained view (and so a less than complete list of a less than core set of features) may be more apt for describing different phenomena.

Take one of the criteria specified as in the logical core of the race concept: HLC1 (visible physical features of a relevant type) –or what Haslanger refers to as 'color'. This is like the arm of the pen-cap of a pen. It is because of the particular feature of having an arm –which is supposed to help pens fit pockets of suits, where pens are normally held- that I can make the particular use of a pen-cap as a hair-clip[33]. Similarly, though HLC1-3 make reference to features of the sort relevant to 'race' –which is supposed to help export the concept correctly, and have it used where it is supposed to be normally used (you can fill in what that use is depending on what you think 'races' are), the fact that the concept can be used to pick out humans' bodies, in specific ways, becomes a very useful fact about it, and a useful property the concept has for a reason not foreseen by those who first started classifying by race; this is a feature of the concept, which we can use in a context where we don't care about race, really, but where we care instead about bodies, and in specific –say- bodily responses to angiotensin-converting enzyme (ACE) inhibitor therapy [a drug for heart disease].

[33] As will become clear in Chapter 3 Section 1, there are other enabling factors to me finding and founding the pen-cap as a hair-clip; which is what I want to hold visible by making reference to a. contexts "space" and "interest" in the found science story, and b. to the right tools for the job of finding and founding.

Take the context of a clinical trial carried out to test an angiotensin-converting enzyme (ACE) inhibitor therapy, say that of Exner et al (2001). Patients are already "trained" to self-identify by 'race' and doctors are "trained" to recognize 'race'. This is a feature "bought" along with our patients like a pen-cap is bought along with a pen[34]. We don't care about figuring out what race really is in that trial. No. We care about the drug we're testing. We care to see who out of the patients with left ventricular dysfunction is worse off after an angiotensin-converting enzyme (ACE) inhibitor therapy. This is like me saying that I don't care about what the pen-cap "really" is; I just want to hold-up my hair. What will I use? It's convenient that the pen-cap comes along with the pen, and that I usually need to hold up my hair while I'm writing. But that doesn't mandate the use of a pen-cap as a hair-clip.

In the context of a trial, we happen to use 'race' (along with other features) to divide patients into groups. We can do this because features of the concept of race allow us to so do: this would probably be criterion HLC1 or 'color' if we're other-identifying race, and perhaps a mixture of criteria in the case of self-identification. But really it is not people that we're interested in sorting here in the context of this trial. What we care to sort are measures of health outcomes of these people.

We obtain a set of measurements –data- on these patients, using all sorts of variables and instruments. This data will include values for variables like 'treatment assignment', 'sex', 'left ventricular ejection fraction', or 'age'. In the case where we think one of these measures is causally responsible for the outcome we are interested in, the factor is controlled for: people with matching values for this characteristic are matched across treatment and control groups to ensure there is no imbalance in our data. This can be done in turn with different variables. The results are then statistically analyzed to see whether variations in these health outcomes show significant patterns of correlation with each other. But notice that through all this we are no longer talking about people, or the shapes of their noses, or how harshly they grew up, or any of that…

Is This Really Still A Concept of Race?

[34] Not all pens come with the "right" sort of pen-cap for my use; and not all patients come with the "right" visible physical features for medical use.

The race concept used to sort health outcomes is a found epidemiological variable. It is found as a census category and founded as an epidemiological category by being repeatedly used to sort measurable health outcomes into groups. This is like saying that a real capacity the pen-cap has to clip on things allows it to clip on more than pockets of suits: it allows it to clip on my hair.

If it happens (as it does) that the variable ‘race’ picks up some interesting pattern in how health outcomes after ACE inhibitor therapy are distributed this will be because of a feature of ‘race’ but it need not be a feature that we have a Real story for: just because this result can be obtained, does not mean that race was “designed” to provide it. Just because a pen-cap is a found hair-clip in the context of my use, doesn’t mean it is Really a hair-clip. Of course, in epidemiological contexts of use of a race concept there is good knowledge that may allow us to infer a causal story for why race thus performs. But the point is that this is extra work. We’re not carving “ACE inhibitor therapy health outcomes” into their joints by noticing their racial stratification; we’re just dressing them up with what outfits we have available fit and seeing how the “look” fares in a broader scientific context.

This per se is not problematic. But being confused about what we’re doing is. In the context of many epidemiological studies ‘race’ is thought of interchangeably as a social confounder and a biological category: when a study controls for 'race'-associated differences in socio-economic status (SES) and education, ‘race’ is taken to be the sort of factor that influences one’s SES. But when used to explain residual risk after such confounders are controlled for as attributed to inherited, genetic susceptibilities, race is used as a class tracking biological inherited differences. And this could be in the context of the very same study; even if one may trace these two different understandings of ‘race’ to different disciplinary frames or practices used for interpreting ‘racial’ data. These race concepts one might call a “found” ‘social status’ concept or a “found” population concept, respectively [I say ‘might’ because a full specification requires a description of the use the concept is put to].

So, more generally,

1. 'Race' can be a found social category if its core criteria are rearticulated to fit a sociorace notion.
2. ‘Race’ can be a found biological category if its core criteria are rearticulated to fit a biorace notion.

This is a heuristic I propose for understanding how a concept that was or was taken to be a ‘race’ concept initially may come to function as say a ‘social status’ concept or a biological ‘population’ concept, ultimately.

This attitude and the admission of contingency in how our concepts are shaped gives me a way out of the seeming paradox that a race concept is not properly a concept of race. I propose that what I called ‘sociorace’ and ‘biorace’ concepts are in effect “founded” concepts: ones found to be social and biological concepts respectively; they function primarily not as concepts of race but rather as either socially or biologically respectable concepts.

1.5 Distinctions Between Race Terms, Concepts of Race and Race Concepts

Before proceeding in this discussion allow me to make the following clarifications.

First, I assume that:

1. There is the term “race” or terms such as “black”, “white” etc.; these I put inside double quotes, or call “race terms”, or “the terms race/ethnicity”
2. There are concepts of ‘race’, or race concepts like ‘socialrace’; these I put inside single quotes, or call them “concepts of race” or more generally “race concepts”.
3. I speak --rather ambiguously-- of “race concepts”; these are either 1. concepts tagged by usage of the word “race” which are founded concepts and so, as I will suggest, properly distinct from “the” concept of race and possibly more akin to a concept of X (where X is specified by the use found for ‘race’ in a target context C) or 2. concepts of race.

4. This dissertation is not about what groups that the ordinary concept of race purports to refer to; but when referring to these, races, I don't use quotes.

I take specifications such as "the concept of race" or "the term race" even when race is not in quotes to not be ambiguous.

This dissertation examines how race concepts are used in biomedical practice; it does not offer a general view on what concepts are, beyond arguing that there are some concepts, such as 'race', that can be found and founded in different contexts of use following different interests, specifically such as the concept of race, as found and founded in biomedical contexts following particular biomedical interests.

Any remarks about concepts that I make in the thesis rely on the understanding of

- 'A's concept of X' as: what A means when they use the term "X";
- 'A's manifest concept of X' as: what definition A gives for X in the context C; and
- 'A's operative concept of X' as: what use A puts X to in the context C.

I don't take these assumptions to be trivial but nor do I think they are atypical assumptions to have in a philosophical discussion of concepts.

I will also use the past participle of the verb 'to find', 'found', as an adjective as follows:

An x is a ***found y*** when x is used as if it were a y.

This is equivalent to saying that x is founded in a context of y-use.

With respect to a common concept:

A common concept is a found 'scientific concept' when the common concept is put to the same uses as the scientific concept would be put to. This is equivalent to saying that the common concept is founded in the context where the scientific concept would be used, or founded in a scientific context.

I proceed to explain my reasoning behind the latter definitions in more detail.

Chapter 2
SYMPTOMS

Some Positions on the Use of 'Race' in Biomedical and Genetics Research

The accounts I examine conceive of race in two ways: a. as a social kind and b. as a biological kind or what I have defined as 'sociorace' and 'biorace' notions of race respectively. This can be problematic in practice as biorace and sociorace notions are often conflated by indiscriminate use of the term "race" to tag them.

Should "race" categories be used to approximate "medically interesting human genetic variation"?

I review the answers of Michael Root, Abdallah Daar and Peter Singer, and Ian Hacking's in that order[35]. Only Michael Root asks and answers this exact question. The others' answers are inferred from how they use the categories and the recommendations they make on related questions.

The first aim of the chapter is to examine the use of the term "race" in these views and organize the concepts "race" refers to. The target race notion of Michael Root is a sociorace notion and most likely 'socialrace'. Bioethicists Abdallah Daar and Peter Singer are working with a biorace notion. Hacking answers the normative question relative to the context the question is asked in. Where Root and Daar and Singer base their normative recommendations on assumptions about what races *are* Hacking's normative recommendations regard the behavior of the variable "race" within particular contexts[36].

[35] More recent work by John Dupre (ms) and Philip Kitcher (2007) is not part of my intended study though these views could be subject to a similar analysis. Dupre's view comes close to that of Root and Kitcher's more recent view comes close to that of Hacking.
[36] This seems to be in the spirit too of biomedical discourse on the topic (cf. Appendix 2).

So this first project of the chapter is a “conceptual analytical” one. As such it does not follow standard philosophy of science discourse: It does not address the arguments these discussants make using said race notions. But there is one more aim in the chapter. And this is a normative one: distinguishing between the two types of race notion I call “biorace” and “sociorace” notions and making the distinction manifest would improve the quality of arguments raised to address the question orienting this discussion. Where philosophical content relates to the project of found science, I outline this relation.

To argue for this target notion Root is operating with both biorace and sociorace notions and not manifestly so and so contrary to what Root claims biorace notions must be added to the sociorace notion he sets as a target one. Daar and Singer seem to underestimate the salience of sociorace race notions for structuring and delivering the results of biomedical research obtained along the lines of bioraces. And Hacking’s account makes visible a possibility which however Hacking does not articulate. The use of the variable “race” can be evaluated coherently without relying on a broader metaphysical/ontological assumption about what races are.

So, answers to the normative question vary. Some answers put up STOP [or GO] signs about using “race” in medicine. Some focus on the variable “race” and talk about all the –mostly wonderful– things *it* can do. This dual role “race” can play and its foundations are explored in Chapter 3 DIAGNOSIS. Both when one type of notion is favored over another as “the one” to use in answering this normative question, and when the distinction is made but not made manifest, we get bad answers to the normative question orienting this debate: an indication that for normative questions about “race” to be settled inter-disciplinary thinking is needed.

2.1 Michael Root on the Usefulness of ‘Race’ for Epidemiological Research and the Usefulness of ‘Ancestry’ for Biomedical Genetics Research

My treatment of Michael Root’s work takes the lion’s share of this chapter. Root has written extensively on methodological problems regarding the use of ‘race’ variables in the social

and biomedical sciences[37]. Plus his thesis on the issue happens to be widely endorsed in fields outside philosophy of science[38] even if in most instances it is not attributed to his specific authorship.

First, I clarify the theses of Root. Root argues that race can have a place in science without having a place in biology: it is a real social kind (2000). He proposes that race should be primarily used to study the effects of racism on people's health and that a more "social" understandings of race would be most useful for biomedical research of that sort (2001). Finally, he claims that race is a bad proxy to use to track any medically interesting human genetic variation and proposes that ancestry would be a better category to use for these purposes (2003).

What may be isolated as Root's manifest concept of race is a sociorace notion and specifically socialrace; but his operative concept of race is in some instances a biorace notion. His double usage contradicts his normative suggestion that we more for a more "social" understanding of race in biomedicine.

More generally, Root's argument suffers from overestimating the importance that considerations of "reality" have in structuring science practice. Race may be a real social kind; and this fact may to Root justify why it should be studied in science. But it does not specify a condition of ontological import as to why race *can* be studied in science nor is it an accurate description of how race is *being* studied in biomedical practice. The variable 'race' can be coherently thought to capture either social or biological variation across populations studied and Root himself operates with such notions. This is a fact that found science begs to understand.

2.1.1 Race Can Be Used in Biomedicine: It is a Real Social Category

I begin with Root's argument in "How we divide the world" for why race can have a place in science without having one in biology (Root 2000).

Root says his aim is:

[37] See Root (2000), Root (2001), Root (2003), Root (2005), Root (2006), Root (ms)
[38] See for example Jones (2001) or Sankar et al (2004), Duster (2004)

> to explain how a social category, a kind like Race, can be both invented and real, and how the distinction between blacks and whites can have a place in science without having one in biology (S628-S629).

Notice three theses here: 1. Race is a social category, 2. a more general thesis: a social category can be "both invented and real" and 3. the more specific claim: 'race' can have a place in science without having one in biology. The first thesis is here taken more or less as an assumption. And though it is not Root premises the specific verdict on the aptness of race categories for scientific research (3) on the more general thesis regarding the reality of social kinds like race (2). He takes the type of reality that an "invented" category has to legitimate the particular use of this invented kind (race) in science.

Should "race" categories be used to approximate "medically interesting human variation"?

Root's explicit aim is not to argue that race *should* have a place in biomedical research but rather to argue that race *can legitimately* have a place in science. However his argument imposes too high a demand on scientific legitimacy: i.e. its "reality" as opposed to its "inventedness". Even though race is invented, race is real and as it is real, we can be using it (i.e. measuring it) in science. I first examine what concepts drive Root's argument.

2.1.1.1 Root's Operative Race Concept: Socialrace

Root claims that race can be used to get scientific generalizations without being a biological category. The kinds of generalizations he means are epidemiological and social science results obtained along U.S. race categories like: "blacks are seven times more likely to die of tuberculosis than whites, three times more likely to die of HIV-AIDS than whites and two times more likely than whites to die of diabetes"; or "60% of all female headed households are black while only 18% are white" and "the inmate population is 50% black and only 35% white" (S629). These epidemiological generalizations do not imply that races are biological says Root. Rather, race affects health via "affecting" income, housing, health care and stress ("being black in the U.S. today is stressful" and stress "suppresses the immune system") (S629). So, "the diseases are biological but the racial differences are not"(S629).

Further, it is racism that is often at the root of why race affects social conditions which in turn affect health outcomes:

> racial differences in social or economic status or in rates of disease have a common cause; they arise from racial discrimination in employment, housing, education, health care, and the criminal justice system. That is, *much of* the variance between the races in socioeconomic standing, as well as health and disease, is explained by past or present acts of discrimination based on race (S629)[39]. [Emphasis added]

What is Root's concept of race here?

Root's operative notion of race is a sociorace notion: a notion that distinguishes between human groups on the basis that they are taken to be 'races' in a particular (U.S.) social context. Root adds that our understanding of what is going on is obscured because though race in the U.S. is not biological, it is "biologically rather than culturally *transmitted"* (S634) [emphasis added] meaning that biological offspring of members of race R are taken to be members of the same race, while adopted children of members of race R need not be members of R (S634) [40]. So, the first order race notion in this sociorace definition is a biorace notion.

Further, Root claims that race is "biologically transmitted" on the basis of a false assumption: that there is a gene for race. This false assumption when supplemented with the correct assumption that parents transmit their genes to their children implies that race is genetically transmitted from parents to children. Hence this biorace notion is a false one: the mistaken notion on the basis of which we think that race is biologically transmitted.

Root understands race to be a social kind that picks out differences on the basis of a belief about their biology. Root could be referring to what Hardimon defines as the ordinary concept of race or to the racialist race concept as the notion that assumes –mistakenly– that there is a genetic basis for racial differences. But given the further understanding of "race" as a category along which discrimination has historically operated narrows it down to a *racialist* race

[39] Root does not attribute *all* variation in health and disease measured across races to discrimination based on race, but he also doesn't specify what causes residual differences may be put down to.
[40] Haslanger would disagree with this last claim. She thinks that her black adopted children are more white than other black children and that she is more black than other whites. Cf Haslanger (2004). Note also that Kitcher (1999) defines the transmission of race in a similar way as Root does here.

concept. Given that a sociorace with racialist race at its first-order race concept was how Hardimon defined socialrace, one may thus conclude that Root understands race as *socialrace*

2.1.1.2 Race: a Real Social Category

The conundrum of a non-biological category having biologically (epidemiologically) measurable effects seems settled, and along the way the point made that these are significant generalizations we are talking about –and possibly very useful ones. But Root wants to further justify the legitimacy of these categories so he argues that they are real.

Root contrasts the capacity that race has to enter into generalizations with categories that "are simply myths or illusions" as would be, for example, the categories of sexual deviance "zoophiles" and "zooerasts" that psychologist William Krafft-Ebbing used in his research (S630): Race is not "nature's category": "race was unearthed rather than imposed by the social and biomedical sciences for once we divided ourselves by race, there were races for a science to observe and investigate"(S630). So, Root stands counter to the possibility that racial categories are part of self-fulfilling social and medical science research that some constructionist accounts may favor.

But this is not the whole story. "Were realism simply about generalizations race would be real", he claims (S631). However "[R]ace does not travel" (S631). Root puts this down to a general difference between natural and social kinds: Terms like 'race' enter into generalizations that are local, because racial categories are locally, socially defined, contrary to biological classifications that are independent of the method of identification and independent of the locale where the identification is made (S631). Root claims that "[A] group of people must divide themselves by race but not blood type in order to have one" (S632)[41].

"How is race <u>real</u> in the U.S. today?" he asks (S633) [Emphasis added].

[41] This is not a view shared by a robust [if not main] stream of science studies work. Root's view neglects work that emphasizes the social and historical production of categories like blood type, which Root posits are "had" independent of a method of identification, contrary to social categories.

> With classification in the natural science, real categories sort individuals on the basis of what they are (by nature), while in the social, they sort on the basis of how, according to the subjects, they ought to be (S633). [Emphasis added]

In agreement with John Elster Root distinguishes between the source of order in nature as 'constant conjunction' and the source of order in society as 'norms and regulations'. He claims it is this difference in the type of normativity bounding social as opposed to natural kinds which is responsible for why real social categories unlike natural categories can be observed in their breach: "kinds of elements or compounds in chemistry simply are, there is no way that a drop of water or bit of hydrogen ought to be", he says, whereas blacks didn't have to make up only 1.3% of US lawyers in 1970 (S633).

How are we then to understand "reality" in the case of social categories? Root defines a real *natural* category (RNC for short) as follows:

> (RNC) A naturally occurring category K is real if and only if K makes extrapolations of many discoverable traits possible across all K things (S633).

The notion of a natural law or that of "naturally sourced" normativity is hidden in this definition; I take it it is embedded in the notion Root has of a "discoverable" trait where a discoverable trait is understood to be discoverable because it is (really) there [not because our instruments are misfiring or such][42]. This definition may leave some clarity to be desired but at least it is there.

Root gives no definition for what would be a real *social* category. He claims that

> Where K is a social category, extrapolation across all instances is not possible, but normalization is, for if real, K prescribes how all K things ought to be rather than how they are (S633).

Strictly speaking a category doesn't prescribe anything. It only states and at best describes. The normative force Root understands a category to have involves its application: the "ought" in the definition points to Laws of Man (Society, etc.) instead of Laws of Nature as the target generalizations into which 'real' social science variables must enter.

Following Root's definition above for natural categories, his assertions regarding the capacity of real social kinds to track social normativity and his distinguishing between real and

[42]As before, Root shows more faith in the epistemic virtuousness of our natural/social science methods than a standard science studies reading would.

illusionary categories on the basis of empirical evidence that the categories enter into significant (though local) generalizations, it would be safe to attribute the following definition of a real social category (RSC for short) to Root:

> (RSC) A social category K is real (if and) only if K makes extrapolations of what discernible traits things K ought to have according to subjects and/or norms in the context in question (and K enters into significant though local generalizations).

Note that if we follow RNC and RSC a naturally occurring K need not be a real social category and vice versa.

This sums up Root's argument for why race can have a place in science without having one in biology:

a. Race enters into local generalizations, contrary to categories that are myths or illusions.

b. Race enters only into local generalizations, contrary to biological categories or natural kinds.

c. Race enters into local generalizations by force of social normativity or convention, as opposed to a force of Nature.

i, j, k imply that

d. Race is a real social kind.

But his stated aim was to show that: Race can have a place in science without having one in biology;

> how a social category, a kind like Race, can be both invented and real, and how the distinction between blacks and whites can have a place in science without having one in biology (S628-S629).

What he did argue was that race was a real social kind. How does that imply that race *can* have a place in science?

One must assume Root is not talking about possibility here but rather legitimacy. And a positive answer to the question of the reality of race (whether social or biological) is taken by

Root as a minimum condition for the legitimate use of race categories in science. Otherwise his whole argument is beside the point.

But of course this wording is wrong. Properly speaking the reality of a phenomenon is not a minimum condition for its legitimate use in science. Though I will suggest that its "found" reality (or its founded fictitiousness) might be.

2.1.1.3 Some Worries About Root's Distinctions

Root's repeated distinguishing of natural from social science or social kinds from natural kinds is made on some false premises. Root claims for example that scientific classifications are not normative whereas social classifications are; that there is no debate on conserving natural classifications (S633) whereas social kinds are mutable (S638); that natural kinds enter into generalizations of unlimited scope whereas social kinds enter only local generalizations (S632). These assumptions neglect work in the history and philosophy of science which investigates the epistemological and historical bases for socially normative and historically mutable natural science classifications.

Michael Root grounds the legitimacy of using 'race' in science on its (social) reality. But what about its "inventedness"? Found science is compatible with Root's claim that there are more and less important invented categories for science but it doesn't do so via talk of their reality.

To Root's claim that: "But society's categories are different from nature's not in being less real but in being man made" (S633) found science counters that: Nature's categories seem less mutable because people talk back more than inanimate matter, not because they are real in a different way. To Root's claim that: "there is no way that a drop of water or bit of hydrogen ought to be" (S633) found science counters that there are ways in which drops of water "ought to be"; these are ways corresponding to the ontological or metaphysical demands imposed on the drops/or their concepts by the scientific context(s) in which the concept 'water' is founded.

Treating entities symbolically imposes a normative demand on "naturally" or otherwise occurring entities irrespective of what laws of nature they may be subject to; they too are to be

used as labeled –but need not be so used. Whether or not natural entities are subject to laws of general scope, they too, like social categories can escape their labels and be used in ways thereupon unforeseen, and possibly ones which contravene their label.

And added to the metaphysical dimensions along which natural science thinks found science adds explicitly that of “interests” (a notion which needs further analysis but at least indicates which direction this analysis should follow).

Conclusion

Dealing with ordinary phenomena in social or other scientific systems involves embedding our ordinary concepts of these within social or other scientific systems. This is no easy task. It involves rearticulating our concepts in the specific terms of the context in which we use them. This is a normative enterprise and one that may result in such drastic changes to an original concept that extra work is needed to map the(se) founded concept(s) back to what they were supposed to describe. Any student of physics should have more than one concept tagged by the word “mass”; what is impressive is that physics students should have some common notions of mass and what is further impressive is that they may often express these notions in the same way –but what is sad is that they may not be able to explain it to a lay person.

As such, it seems that a “naturally” occurring kind would be found and founded in science by acquiring the type of reality Root attributes to social kinds: its concept would be bound to the norms that scientific concepts “ought to have according to subjects and/or norms in the *scientific* context in question”. There are accounts of what these norms might or should be; much like there are accounts for what our social and political norms might or should be. They could be dictated by what Lakatos calls a research program or Kuhn a normal science paradigm.

Finding and founding concepts occurs within contexts of space and interests that can and do change historically. This plays out as both a strength and weakness of scientific enquiry; but it is what it is. Other than that I agree with Root’s claim that race can be both real and invented. But

urge him to hold him to his word: the "inventedness" as well as the "reality" of race matters for how we may use it in science.

2.1.2 The Usefulness of Race in Epidemiology

Root examines the specific use of race categories in biomedical epidemiological U.S. based research in "The Problem of Race in Medicine" (Root 2001). The central thesis of the article is that race "should not be dropped as a variable in epidemiology or medical research but that its use should be limited to studies of the impact of racial discrimination on health" (21). This is the thesis I referred to earlier as one that is shared by researchers working on this topic outside philosophy. This is a thesis examined more fully in a later section. But Root also thinks we should move towards a social understanding of race. This is a position understandable given the benefits of understanding race as sociorace in an epidemiological setting which Root makes clear, but it is mistaken. It contradicts Root's first thesis.

The impact of racial discrimination on health cannot be studied in medicine using only a sociorace notion of race –or so I claim.

2.1.2.1 Root's Manifest Target Race Concept: Sociorace

Should "race" categories be used to approximate "medically interesting human variation"?

Root argues that given that we use "race" categories in biomedicine and notice patterns in health outcomes by this "race", "race" should have some role in explaining these differences (27). How should race be understood though, in this program?

Root claims that:

> The best way for medicine to achieve President Clinton's goal of eliminating racial differences in rates of disease[43] is to move away from a biological conception of race and toward a more nuanced social one (35).

[43] Here I take it that Root is referring to the goals set by the NIH directive Healthy People 2010 and specifically to REACH 2010 (Racial and Ethnic Approaches to Community Health) voted in by the Clinton administration.

I don't disagree with Root's normative suggestion; that is, not in full. I certainly think that a sociorace notion is useful for epidemiological work and that Root gives a strong argument for this thesis. This would be an argument for the usefulness of sociorace notions and in particular socialrace for biomedical research and a thesis which is examined separately in section 2.1.4. But I also think that biorace notions are part of the conceptual tools epidemiologists must use. Moving towards a socially nuanced notion is defined negatively by Root as moving away from a biological conception of race; but this assumes a biological conception has to be used; if only to move away from it.

2.1.2.2 Root's Operative Target Concepts

Root's particular arguments for how a biorace notion may lead researchers astray are not enough to dismiss the general use of biorace notions in biomedicine. One cannot base such a general thesis on a few examples. So I will not examine these particular examples here.

Instead, assume that Root is right. Assume that sociorace notions are what we should aim for in biomedicine. Could we make do with only sociorace notions of race in medicine, given current biomedical theory? And why would we want to?

I examine one of the examples which Root uses to support his general thesis. Root argues that attributing what residual epidemiological risk is not controlled by social status confounds to genetic causes is usually mistaken. And he uses the case of differences in "black" and "white" U.S. infants' birth-weight as one of the examples that are telling to this effect (28-29). However he takes such studies to imply that it is therefore useful to move towards a social understanding of race in biomedical epidemiological research. And this is a step that I argue he cannot take –and in fact that his own example shows that he cannot take this step.

Root's thesis is that:

> To conclude that an unexplained variance is due to black or white genes, an epidemiologist must have independent reason to believe that there are genetic difference between the races. But there are no such reasons. Rather, the direct evidence, the fact that there is more genetic variation within than between the races, indicates that the unexplained differences in average birth weight are due to hidden

> (social) variables (e.g. intergenerational differences) rather than any race-linked genes (28-29).

To dismiss genetic explanations of low birth weight for black as compared to white infants, Root cites a study which compared birth weights of infants of U.S. born blacks with birth weights of infants of African born blacks and showed average birth weights of infants of U.S.-born white mothers and African-born black mothers are closer to each other than to the weight of infants of U.S.-born black mothers (28). Root says that "assuming that African born black and U.S. born black mothers are equally black" the study suggests differences observed in 'black' and 'white' infants' birth weight is due to social or economic factors that differ between the U.S. and African context. It is not specified whether the social and economic backgrounds of selected subjects were matched. I assume that, as is most common, African-born blacks were selected according to features that are biological in what Hardimon (ms1, 14) would call a basic sense.

But Root doesn't notice that the assumption of patients being equally black would be nonsense were we to take his frame as normatively as offered. The study's design cannot be expressed, let alone its conclusions (which Root enlists) justified without making first-order use of a biorace notion of race, even if this is to dismiss it.

Were we to use only sociorace notions of race to give a causal explanation of these results we would need one socialrace which makes the U.S. and African groups "equally black" and one that doesn't so that the differences in their health outcomes are explained. The second sociorace notion, the difference maker, is presumably U.S. socialrace -or so Root wants to argue: having a mother of the black U.S. socialrace as opposed to a mother of a non-black U.S. socialrace, makes infants more susceptible to low birth weight. But what would the other sociorace be? If 'black' is defined as the black U.S. socialrace and given that the U.S. social context is different from African social contexts, the definition wouldn't pick out anything in Africa. If we use a socialrace notion defined in an "African" context (which is ill-defined in social science terms) we need not pick out the same "black" people there as in the U.S. as what groups are taken to be racialist races in each context also most likely differs, plus we may not in that case legitimately posit that the impoverished birth weights of American-born blacks are due

to differences due to differential social pressures as presumably groups would be matched in these respects across continents. We might opt for a global sociorace notion: the 'race' subjects participating in the study would be taken to be in a global social context. This seems again problematic. Even if we don't go for a global socialrace notion, so that we avoid positing that there are globally endorsed racialist race notions, a global sociorace would have to be defined as vaguely or as mistakenly as a "global society" currently is.

So what is the notion of race we need to use to articulate the assumption that these subjects are "equally black"?

What about thinking counterfactually? Identify groups outside the U.S. using a U.S. sociorace notion, say socialrace, as follows: call "black" those who were they in the U.S. would be taken to be members of the "black" U.S. socialrace. Counterfactual reasoning is attractive because once we counter all the facts we have a choice on what facts to assume not countered. For example, say the sociorace concept we use is socialrace. Then we may choose which bits of a U.S. socialrace notion, which one of the criteria defining it, will function to pick out subjects in the African context. Instead of wondering whether the African social context conceives these subjects as members of racialist races and subjects them to racist treatment we can use criteria specified by the first-order notion of racialist race, and even more precisely, go for phenotypic characters specified in the logical core of the notion. And we can ask: do African-born subjects have the "physical features of a relevant kind" that would classify them as "black" in the U.S. social context? Or, do African-born subjects have ancestral origins that would classify them as "black" in the U.S. social context? Etc. If so, then call these subjects "equally black" to U.S. born blacks.

But this type of reasoning operates in biological terms; it references traits that are biological –it makes use of what I call a biorace notion, a found biological notion. This is the type of notion Root says "[medicine] should move away from" (35).

Though Root's example cannot prove his general thesis a look at how Root's argument fails in this particular case points to a deeper problem with his logic.

2.1.2.3 Operative Race Concepts: Sociorace and Biorace

I am not sure why Root is not bothered by this. I am not sure why he takes conclusions reached on the basis of using bioraces to question socialraces' biological reality to disqualify biorace from entering biomedical epidemiological discourse as "really" real. But it seems to me that this is very likely how scientists pick out their subjects; and how they define the dimension on which these subjects are assumed to be "equally black". And it seems to me that there is no obvious way to argue to such conclusions using only sociorace notions.

So, what is medicine to do?

What is medicine to do when, further, all sorts of biorace notions are still competing for explanatory power in answering to the regularities recorded using "race" variables?

Take the discussed example again. One can think of typological biorace notions which can be used to frame hypotheses about how it is the racial degeneration of U.S. African black stock through increasing rates of racial intermarriage in the U.S. that explains the differences in birth weight measured in the study that Root cites, saying nothing about socioeconomic explanatory factors nor letting go of a racialist race notion. A racialist race notion would perhaps have more trouble explaining cases where the health outcome said to differ across races is, say something like 'death by homicide' an outcome one more plausibly caused by socio-economic status differences across races than any genetic difference. But still, things are tricky. As the genetic bases for different mental or behavioral traits continue to be investigated one can imagine that a racialist race notion could be used to explain why behavioral traits that cause homicidal tendencies are differently distributed across racial groups. Though it is infuriating to consider such hypotheses, these answers are possible answers to the questions of interest and these are answers that need to be considered and dismissed on the basis of concrete scientific evidence targeted to address them.

Conclusion

Root claims:

> The problem of race in medicine is not that health statistics are stratified by race but how they are. Race is bad for medicine if the racial categories medicine employs fail to match the complex ways that race functions in the United States and, in particular, how race functions to harm or exclude (35).

Root stresses the complex ways in which 'race' functions in the U.S. But would it not be part of the complexity of race that it is actually biological if even in a basic sense? How would a social conception deal with that type of complexity?

Instead of ceasing to investigate the logical, epistemological and biological foundations of ordinary biorace notions and allowing possibly mistaken biorace notions to proliferate I think that we must, as we have so far, use biorace notions to articulate biomedical and biological thinking on these questions.

More generally, what I hope I have suggested through this treatment of Root's argument is this: there are more levels of "reality" on which these seemingly straightforward questions about the use and usefulness of "race" hinge; more than the ones Root deems relevant for answering the practical questions that concern us in medicine. By taking the analysis down from the level of what is manifestly the concept endorsed by Root (socialrace) to what about such a notion would do the work required of it (its biorace core) I aim to challenge his assumption that science can or should proceed only when dealing with the real[44].

2.1.3 'Race' Should not Be Used to Approximate Genetic Variation

Root's discussion of "The Use of Race in Medicine as a Proxy for Genetic Differences" was my introduction to this topic (Root 2003). Its thesis initially seemed very attractive; it seemed common sense. But its very slippage into common sense is why it must be rejected. Understanding and using races as proxies in a clinical setting involves specialized measurement and testing that needs to be responded to rather than overlooked by common sense understandings.

[44] One wonders whether it is possible to assert x is y without an understanding of not-y; and whether, what I'm saying here is as trivial and as non-trivial, as that…

What is Root's attractive thesis? Root argues that race should not be used as a proxy for genetic variation because ancestry can be a better proxy than race and because using race as a proxy for genetic variation is ethically pernicious. So his thesis relies on two premises: 1. ancestry is a better proxy for genetic variation than race, 2. using race as a proxy for genetic variation is harmful. I question both premises in this dissertation, in this chapter and also in the next one, as long as we can understand 'race' in a more nuanced way as sociorace and biorace. I argue that the first premise rests on problematic notions of a 'proxy', 'race' and 'ancestry'. Second Root's ethical worries arise partly from general methodological problems to do with applying statistical knowledge in a clinical setting buffering any force his argument may have in the case of race. Further, what risks pertain specifically to using "race" as a biomedical variable are not particular to using race in genetics, even if they are higher there; therefore Root's argument is asymmetric and limp in that sense, unless Root is prepared to say that we should not be using "race" in epidemiology either for similar reasons.

2.1.3.1 Three Theses, Three Concepts: 'Race', 'Proxy', 'Ancestry'

In this article Root answers explicitly the question I chose to orient this discussion by:

Should "race" categories be used to approximate medically interesting human genetic variation?

Root makes a negative and a positive proposal. He answers this question in the negative: what he calls "a proper understanding of race" questions using race as "a marker of any medically relevant genetic trait" (1173). This, he thinks, is a road we shouldn't but also one that we need not be going down on. Instead of 'race' we can use 'genetic ancestry' which he argues is a better proxy for such variation. So in sum: "race should not be used as a proxy since genetic ancestry has more predictive power and, unlike race, does not require doctors to draw a color line or decide which of their patients are black and which are white" (1181).

How does Root argue for these theses?

His argument raises three concerns regarding the use of race to approximate medically interesting genetic variation: 1. an ontological one, 2. a methodological one and 3. an ethical one that can be correspondingly tagged by three key notions on whose understanding they respectively rely: 1. the notion of *race*, 2. the notion of a *proxy*, 3. the notion of *ancestry*. I organize them thus and discuss them in turn.

2.1.3.2 What Race Is: Understanding 'Race' as 'Socialrace'

Root's first challenge involves the ontological status of *race*: he says that race is a social status variable and not a meaningful biological class. As such it cannot reasonably be expected to track medically interesting human genetic variation, and certainly it should not be expected fare better than the notion of 'genetic ancestry' –the subject of his positive proposal– which as an added benefit doesn't come loaded with a history of abuse by racist ideologies.

This discussion takes place largely in the first part of his paper under the section 'Race as a Category in Epidemiology' and reproduces earlier arguments on why race can be a useful category for medicine even if it is not a biological category.

As in earlier articles Root lists epidemiological statistics that show that people classified as 'black' in the U.S. tend to have poorer health profiles than those classified as 'white' (1174)[45] and juxtaposes the apparent usefulness of race in capturing significant generalizations with the fact that "most biologists" today oppose all three tenets on which race was thought to be "biological race".

> (A) a conjunction of physical characteristics divide the races; (B) these characteristics are heritable and express genetic differences between the races; and (C) the genetic differences are concordant and result from differences in descent - that, at some point the races were reproductively isolated, differences developed, and the differences have been inherited[46] (1174).

[45]These figures replicate those cited in Root's earlier work; Root (2000); Root (2001).
[46] The three criteria listed in the earlier work of Root (Root 2001, p 23) as (a)-(c) are almost the same as these. These new ones are slightly stricter as criterion (C) is stricter than (c): (C) lists posited genetic differences as concordant, i.e. as varying in dependent ways, where (c) just assumed that the genetic differences were inherited. This modification is important as Root goes on to challenge the correctness of the "biological race" conception on this point.

"Biological race" is a biorace notion: It is a concept that distinguishes groups of humans that were taken to be 'races' in the context of biomedical science. And it is a notion that attributes to these groups characteristics of a biological type. But it is not any biorace notion. Notice that points B and C makes this a typological race concept. This, Root repeats, is not a notion that biologists agree with any more:

> Today, most biologists oppose all three tenets. They allow that biological differences between populations customarily sorted as separate races are at best statistical; the populations, if biologically different at all, are differentiated only by average frequencies of a few polymorphic genes. In addition, the differences are not concordant; the differences between populations with respect to one gene vary independently of any differences in another, and, as a result, there is no cluster of genes possessed by all and only individuals customarily sorted at some site as members of the same racial group (1174).

Each of the points he makes above negates the premises of the biological race notion: contrary to claim (B) variations in the average frequencies of genes' expressions are not variations in the having or not of particular genes: Root calls these genetic differences between these groups "at best statistical" (which is not calling them non-existent). He notes contrary to (A) and (C) that whatever genetic differences exist across these groups do not vary in conjunction and as a result that contra (C), individuals sorted customarily as members of the same race would not possess some common cluster of genes (1174).

Root adds that the fact about the biological meaninglessness of race is contingent on how mechanisms of inheritance and selection have acted on our species and that "[H]ad our natural histories been [sic] different, we might have been more different than we are" (1174). Root privileges a biological recipe as what could have made 'racial' human differences more important than they are: a recipe for how natural process could have given us "biological races" in the case of humans. "First, isolate a breeding population. Second, wait for some distinctive heritable characteristics to appear. Third, give their conjunction a selective advantage. And fourth, let selection operate for a very long time, but be sure to keep the population isolated" (1174). To the best of our knowledge, he adds, Human evolution did not proceed according to such a recipe. Human populations "have not been geographically isolated for long enough periods of time;

during all of natural history, there has been too much breeding between populations to give us biological races" (1174).

He concludes that though what we call "races" are not biological races this doesn't relegate racial classifications to nonsense. Though this "biological race" is mistaken, it forms the basis on which we customarily identify race and thus treat each other differently by race: thus it tracks effects that racist attitudes may have on health outcomes. Race can be "biologically salient" (1175) .

It becomes visible again that Root's operative biorace and sociorace notions are racialist race and socialrace. His operative sociorace distinguishes human groups on the basis that they are (mistakenly) taken to be (biological) 'races' in the social context of the U.S. and as he points to the assignment causing us to "treat each other differently" he hints that the first order "race" that groups are taken to be in the U.S. legitimates excessively oppressive, racialized power structures and so is a racialist race notion (1175). This supports my earlier conclusion: Root's first-order biorace concept is racialist race and his operative race notion is socialrace, at least in the context of this argument. And my earlier critique stands still.

Root takes this argument to debunk the biological reality of race. And as argued already, he also takes the irreality of this particular biorace notion to disbar the use of this and all other biorace notions in biomedicine. But what he has in fact put up for attack is a particular biorace notion; a typological one and a racialist one. But as argued already biorace notions are needed to study race in biomedicine. As I argued Root would need to be in practice using both biorace and sociorace notions of race, to argue for the 'inappropriateness' of biorace notions for tracking causes of health outcomes. This is understandable given his understanding of race as socialrace: of course socialrace cannot track genetic variation. But some biorace could. So Root's argument is launched on a false assumption: Showing that some racialist biorace fails to work shows that all biorace notions will not.

But this is not Root's focus in this article. This first section of his argument lists only what results he wants to take for granted and use as a springboard for his new suggestion.

2.1.3.3 Root's "Methodological" Concern

The second concern of Root looks to be methodological. Beginning a section titled 'Race as a Proxy' Root sets up a problem:

> While epidemiologists use race as a population variable in their studies of difference in morbidity and mortality, doctors sometimes use race as an individual variable, as a way to classify an individual patient. Doctors use race as an individual variable, for example, when they use race as a proxy for an individual patient's response to a medical treatment or as a proxy for a gene (1175).

Root says race, a "population variable", should not be used in the context of a medical clinic as an "individual variable" i.e. as a descriptor and predictor of individual patients' health outcomes. I say this concern looks to be methodological because in fact this concern reduces to Root's earlier ontological one. There is a gap between his manifest definition of what a "proxy" is and how he uses the term, i.e. his operative notion of a proxy. He assumes that the quality of the relation of 'approximation' is one that relies on causation not on correlation and so that as 'race' is sociorace race cannot "approximate" genetic variation any better than ancestry.

2.1.3.3.1 Root's Manifest Concept of a 'Proxy'

Root reproduces here the premises of his earlier arguments on the usefulness of race (understood as a socialrace) for biomedical research because he takes some of the same premises to support his new thesis: that there is no solid epistemological basis on which to consider race a proxy for human genetic variation –but that give reason to privilege 'ancestry' instead. Why? Because, I think, he is making a mistake.

What is it for something to be a proxy? Root calls variable X a proxy for another variable Y in the following case:

> One variable X is used as a proxy for another Y when X is used in the place of Y to make a particular decision about an individual: Let Y be a variable that is material to an interest I but that cannot be directly measured, and X a variable that can be directly measured but is not material to I but correlates with Y. In that case, X is a proxy for Y if X is used instead of Y in making a decision about the individual in order to further I (1175-1176).

Note that Root is defining two things here. The first is when "one variable X is used as a proxy for another Y"; the other is when "X is a proxy for Y". So, break these up and label them as Root on Using as a Proxy (RUP) and Root on Being a Proxy (RBP):

(RUP) One variable X is used as a proxy for another Y when X is used in the place of Y to make a particular decision about an individual.

(RBP) Let Y be a variable that is material to an interest I but that cannot be directly measured, and X a variable that can be directly measured but is not material to I but correlates with Y. In that case, X is a proxy for Y if X is used instead of Y in making a decision about the individual in order to further I.

Now let's look at what he says. Given variables X, Y and interest I, when is X a proxy for Y? Under the following conditions: first X must be correlated to Y, second X must be directly measurable when Y is not, third X must be immaterial to the interest I we want to further to which Y is material. If all these are the case and fourth X is used instead of Y to make a decision about an individual to further interest I, then X is a proxy for Y.

Is RBP a reasonable definition?

It is too strict: it gives sufficient conditions for X to be a proxy for Y but I don't think these conditions are necessary. This is because this definition does not allow for the possibility of self-approximation, i.e. it is not reflexive. If we follow this definition then X cannot be a proxy for X because of the way the second and third criteria are phrased: X is defined as directly measurable when Y is not and as immaterial to the interest Y is material to. It seems that X should be its own proxy, and so these should only be minimal conditions: Y need not be hard to directly measure, and X need not be immaterial to interest I.

So a better definition would be the following (Efstathiou on Being a Proxy):

(EBP) X is a proxy for Y regarding interest I if and only if X is correlated with Y, X is directly measurable and X can be used to make a decision to further interest I to which Y is material.

How can we make sense of Root's "mistake" in the context of this argument?

2.1.3.3.2 Root's Operative Notion of a Proxy

Here is one way: The excess strictness imposed via RBP on a candidate proxy X for Y in furthering I, compared to EBP could be specifying normative conditions for one of two things: a. for being a *good* proxy, or rather, more broadly, b. for being what *good* proxy-use can be made of. So the logical "stricture" from EBP to RBP can be read as transfiguring the notion of a 'proxy' to one of a 'good proxy' or more broadly or one of a 'good variable to use as a proxy'. I am thinking in terms of merit because though explicitly none of the definitions Root gives specifies when X is *correctly* or aptly used as a proxy for Y, nor when X is a *good* proxy for Y, in effect such specifications are clear in Root's mind and applied when considering whether race should be used as a proxy for medically relevant genetic differences. Further support for this speculation comes once his argument is in full view.

The claim that X should be a proxy for Y only if X is measurable and Y is not, indicates that when Y is measurable something else should happen; perhaps we shouldn't use X as a proxy for it, but rather just measure Y. Indeed: "Proxies are only useful when we lack direct access to the material trait" (1180). This is what Root claims for the case of 'race' as approximating 'genetic ancestry'; he will say that genetic ancestry can be measured –or rather assume this, without much argument- and thus disbar the use of 'race' as a variable because it is not the "best available" one we have (1181). Similarly, when he requires that X not be material to the interest we want to further –to which instead Y is material- he is again setting up a problem he will run into when discussing race and ancestry. If the interest we want to further by using 'race' to approximate 'genetic ancestry' in making a decision about an individual is 'that individual's health', then 'race' is material to an individual's health in the context of the U.S. –for that he argues convincingly. This 'material' [read: 'causal'] connection is then perhaps to be avoided because it can be problematic: the use of race as a proxy can have impacts that are irrespective of how well race correlates with ancestry and this fact about race may in fact stunt the interest we want to further.

An example he gives that would fit this description is the case when an individual, though they are being screened so they can be protected from sickle-cell anemia, ends up being discriminated against because his race is associated mistakenly with an essential biological feature (1177). And substituting the 'well-being of the group an individual belongs to' for the interest 'I' we want to further by using 'race' as a proxy for 'ancestry' would bolster an argument that Root makes against using race as a proxy for ancestry: that such a use will hike the risk of racial discrimination for the whole group that may be deemed either prone to some genetic disease, or biologically of essence.

So perhaps Root defines a "good" proxy in RBP. But Root doesn't seem to follow even this definition everywhere!

2.1.3.4 The Final Solution: 'Ancestry'

Even if we take Root's definitions at face value, to be a proxy a variable must *correlate* with what variable is of direct interest but hard to measure; it need not in any way be causally related to that variable. But it seems that in practice Root requires even more from a proxy than his rather strict manifest notion of one.

Root's positive proposal is that ancestry is a *better* proxy than race when measuring genetic traits of medical interest. But the claim is not made on the basis of there being a good correlation let alone a *better* correlation between ancestry and the medical outcomes of interest as one would expect from his definition. No, there is no evidence in the form of empirical results comparing how much more strongly 'ancestry' as opposed to 'race' is correlated to even one genetically based feature of biomedical interest in this article. Rather, Root's positive suggestion on why ancestry would be a better proxy for medically interesting human genetic variation rests on a posited biological causal story.

The only way in which Root's suggested target notion 'genetic ancestry' trumps race in these 'proxy wars' is by referring to groups that Root says are less genetically variegated. Root

makes this point in the context of his discussion of how drug response is measured to vary by race. He says that

> … because there is so much genetic variation within any one racial group, race is a poor way to identify populations for genetic comparison. The human groups that are called different races are not distinct lineages. Genetic differences do exist between human populations in drug response, but the existing racial categories do not capture these differences very well. Ancestry captures the differences better, since there is less genetic variation within groups identified by common genetic ancestry than groups identified by race (1179).

But this is an assertion about what groups identified by ancestry are; it is not an assertion about whether or not these distinctions *do* capture differences better than race nor is it an assertion about why they *must*. It is a metaphysical assumption about what ancestry is that Root uses to bolster this inference, not evidence that speaks to the relevant empirical question: Does identifying human populations using 'genetic ancestry' classifications approximate biomedically interesting genetic variation across members of these groups better than 'race' does? He does not say. This is a problem. Because as evidence from studies like that of Rosenberg (2002) and Tang et al (2005) seems to suggest, and as Daar and Singer next argue, race can be a pretty good proxy for 'genetic ancestry', at least in the case of the U.S. (cf. Chapter 3 Section 3 *Founding 'Self-Identified Race/Ethnicity' in Genetics*).

Further, as Risch et al (2002) argue and as is especially indicated from the study of Tang et al (2005) (Chapter 3 Section 3.3.2.2), 'race', though a category that is not a biological one in origin, is a category that we need to study in biomedical genomics if we are to avoid confounding differences measured across these categories with differences that are genetic in origin.

What if, that is, the differences we measure using 'genetic ancestry' are in fact differences caused by the social status of the people that the categories measure? The burden of proof lies with Root here. If things are just bad, epistemologically speaking, and what races' health outcomes are caused by is a complicated question, then we're better off not pretending to sanitize the objects of our biomedical and genetic interest by re-distributing who would be their members to more kosher categories like 'ancestry'.

The way his argument's logic slips indicates that for Root a good proxy is a category that, sure, is correlated with the phenomenon of interest but also, a category that is or could plausibly be causally (or otherwise, metaphysically) related to the phenomenon of interest. So it is no wonder that using ancestry trumps using race as a proxy for medically interesting genetic variation: ancestry is (or is almost) a result of Nature's recipe, and race which Root understands as sociorace is not. Tracking human genetic variation is not something that a category that is not describing a biological phenomenon can do.

2.1.3.5 The Need to Keep Thinking

Needless to say, I do not think that Root's assumption must be satisfied by a good proxy. I would add to Root's definition that *X* is *aptly* used as a proxy for *Y*, where *Y* is material to interest *I* but hard to measure

1. when *X* is a good proxy for *Y*, i.e. when the correlation between *X* and *Y* is strong (which I gather Root assumes happens when there is a causal story relating X and Y) or

2. when interest *I* is important enough for the individual (or other agent) that even a rough approximation for *Y* by *X* is better than no means of approximating *I* at all.

Given a broader understanding of 'errors' than just a statistical one, one can parse the above as saying that *X* is an apt proxy for *Y* with respect to *I* when the errors of measuring *X* instead of *Y* are negligible in the context of *I*.

Indeed Root brings considerations beyond scientific theory in deciding what is a good and a bad use of race as a proxy for medically interesting genetic variation: even if race were a good proxy for genetics, he says that using race for this job fosters conceptions of race as biologically essential increasing the social and political risks for certain populations[47]. "In considering health care policy what matters is social welfare and race-conscious has less expected social utility than race-blind medicine in today's racial climate in America" (1182).

[47] Richard Cooper et al. (2003), Mary-Claire King and Arno Motulsky (2002), and Troy Duster (2005) also argue along these lines.

Why is the social utility of race-conscious medicine lower than that of race-blind medicine, in this context?

> To stratify health statistics by race is reasonable, as long as employment, housing, income, education, or healthcare are stratified by race; but to use race as a proxy for a gene is bad science, because race and genes vary independently, and bad policy, because the practice helps to sustain a harmful racial ideology (1182).

By "bad science" Root is referring to the dubiousness of the biological ontological status of 'race' which is a worry I responded to earlier (there can be more biorace notions than a racialist typological one). But the second worry is common across the literature on this topic[48], and one reason why race-specific research has met with great controversy in the United States, so let us focus on this here.

History teaches that racist ideologies are harmful and sustaining them is a bad thing. This discussion is a much larger and more complicated one than Root admits. The debate regarding affirmative action –whether it takes the form of affirmative *medical* action or not– is great and beyond the scope of this dissertation. But this is what is involved here: a tradeoff.

There are three assumptions Root makes which he needs to support further. First, the specialness of genetics: Why is continuing to use race to approximate social conditions less pernicious or licentious a practice than using it to approximate genetic variation? Second, the social utility of ancestry: Would racial ideologies be somehow "hit" were we to use 'ancestry' as a proxy for human genetic variation instead of 'race'? Third, the efficiency of "ancestry"-specific genetic scientific research: How sensible is it for policy and for research to introduce this terminological break between neighboring biomedical fields such as social epidemiology and genetic epidemiology? These are questions I cannot fully answer. But they challenge Root's peremptoriness on this issue.

1. <u>The specialness of genetics:</u> Root has argued for the social utility of race-conscious medicine; at least in the context of medicine remedying the effects of racism on health. How is genetics different? Root buys the causal story linking race –understood as socialrace– to social

[48] Richard Cooper et al. (2003), Mary-Claire King and Arno Motulsky (2002), and Troy Duster (2005) also argue along these lines.

conditions impacting health outcomes; he does not buy a causal story linking socialrace or racialist race to genetic variation important for health outcomes (nor should he!). But this doesn't mean that race-specific genetic medicine can have no social utility. Daar and Singer (2005) who think of race as biorace argue that it is at precisely this coarseness of grain, not a finer one, that pharmacogenetics would be ethical (Section 2.2 of this Chapter). Sure, this relies on a different understanding of race than the one Root endorses but it is a plausible one still.

And what about the sustenance to racist ideologies provided by keeping to profile people's "employment, housing, income, education, or healthcare" by race? These practices too lend support to racist ideologies, insofar as these status indicators –even though they are social ones– can be ranked. Typological assumptions regarding people's social status are no less pernicious than assumptions regarding people's biological makeup (cf. Ellison 2005). So how does Root circumvent such problems in the case of this type of race-specific research, and can he not follow the same route to circumvent these in other types of (metaphysically sensible) race-specific research?

2. The social utility of ancestry: There is arguably room for racist overtones to even ancestry-stratified genetic research. Root thinks that ancestry is a term less loaded than 'race' and as such a better variable for people interested in biomedical genomics to use. But this view relies on a particular understanding of ancestry as opposed to race. Ancient geographical ancestry, though different to race, is arguably a notion akin to that of race. If we follow Hardimon, both ancestry (HLC2) and geographical origin (HLC3) are part of the logical core of a notion of race. Though ancestral specifications need not distinguish groups that are phenotypically distinct, the opposite is arguably the case. To each, 'race' there will correspond a cluster of 'ancestries'. Whether we infer from results about 'ancestry' conclusions about 'races' depends on how well we can specify the correspondence between the two but this in principle seems possible. If say racist hiring practices are a reality we face in the U.S. then there is plausibly as much risk for job applicants of say "Sub-Saharan" or "Bantu" origin as ones of "black" race; as the risk can be distributed to what race each ancestry corresponds to.

3. The efficiency of "ancestry"-specific genetic research: What would the use of 'ancestry' as opposed to 'race' involve in practice? As argued later (Chapter 3 Section 3.2.1, and Appendix 3) one of the reasons that race categories have been used in biomedical epidemiological and genomics research is that they were already being used in demography. Should data already obtained according to 'race' be aligned with any new data on 'ancestry', at least in cases where we are interested in studying feedback between environmental and genetic factors affecting one's susceptibilities to disease? If so, then how will the alignment be understood to occur? What about the risk of confounding genetic variations observed under 'ancestry' with the social ones that affect a corresponding 'race' that Risch et al (2002) and Tang et al (2005) imply is non-negligible?

These are all questions Root needs to address before his positive proposal is accepted.

Conclusion

Sociorace is a social, analytic concept. Socialrace, which I argued, is how Root conceives of race, manifestly describes a first order social phenomenon and picks out groups that are identified as racialist races by a society (like the U.S.). Now, if we say that *socialrace* is an adequate proxy for *genetic* variation, we are essentially asserting that a *social* phenomenon adequately captures the organization of invisible [to the lay eye] genetic microstructures. This might very well happen if the actual populations that 'socialrace' is picking out happen to exhibit such genetic heterogeneity (or homogeneity –depending on how the sorting technique followed operates). However, there will be nothing in the *meaning* of the term 'socialrace' that can adequately explain this happenstance; and this won't be the case for a while, at least until there is more theory that explains precisely how environmental and genomic causes of disease 'interact'. Even though the extension of 'socialrace' might include genetically diverse populations, there will be no genetic phenomenon that 'socialrace' can properly describe –assuming that social phenomena do not shape biological entities at the genetic level.

The race concept that appears to be a better candidate for a proxy for genetic heterogeneity would be a biorace concept: perhaps what Hardimon calls the "populationist" race concept, or some other biorace concept out of the many proposed. Biorace notions purport to talk of races as biological phenomena. So, a notion of that type would be a notion appropriate for looking at the question of human genetic variability using terms that are closer to the terms already couching biological theory. Thus, once we adopt this frame, and if only for the purposes of speaking in a coherent scientific vocabulary about genetic causation, there will be no instance where socialrace will be a more appropriate proxy for medically interesting human genetic variability than a biorace notion.

Now, with this conceptual clarification, one might go back to Root's critique and substitute "biorace" and/or "sociorace" where he says "race". This would leave us with a re-articulation of his earlier theses: 1. race is sociorace: it cannot be biorace, 2. socialrace can well-approximate social causes of health outcomes (a thesis discussed in the following section), 3. 'ancestry' is a better proxy for human genetic variation than racialist race and socialrace.

Root's argument suffers because of two deeper problems in the structure of his thinking. First, because Root relies too solidly on a sociorace understanding of race; he underestimates how closely a biorace understanding of 'race' may in fact be related to – and perhaps well-approximate- what he holds as his target biomedical genetics notion: 'ancestry'. Second, his argument suffers because Root though he seems (later too) to talk of 'race' in a pragmatic manner –as varying what we call 'race' to measure the effects of racism, in effect I think he understands science in slightly too purified a way: as testing theory rather than delivering results, as epistemology not technology. Of course these broad claims I cannot prove beyond vague gesturing, but do I think that it is as a result of this type of slant in his view that he sidesteps the possibility that there could be legitimate, independent reasons to study 'race' understood as biorace in biomedical genetics or genomics, where this new [or rather old and newly recovered] 'race' notion does approximate human genetic 'ancestry'.

So, Root's thesis is implied by the following normative claim; a claim that he doesn't

explicitly make but which I think he assumes: Common races should not be used as proxies because they CANNOT approximate medically interesting genetic variation. Why not? Current biological theory tells us that common races are not "biological races": "Though there are heritable differences between us, they do not cluster and do not pick out the classes we call 'races' " (1174).

Fixing his conclusions on what one can do with the category 'race' on his understanding of what races are, Root ends up inferring too much, too quickly and too confidently. Thinking of 'race' as socialrace makes Root ignore possible correlations between 'race/ethnicity' [conceived as biorace] and 'geographic ancestry'. Root proposes that 'geographic ancestry' is a better proxy for genetic variability than 'race', because of the causal story that can plausibly link geographic ancestry to genetic variability. But he is assuming that 'race' is not well correlated with 'geographic ancestry' and as discussed already, Risch et al (2002) and Tang et al (2005) suggest that the two are in fact well correlated for the U.S. population.

To prove that 'x is a bad *proxy* for y' one needs to talk of *correlations*, not causal pathways. The lack of a clear causal pathway linking 'race' to genetic variability in Root's mind is understandable given his understanding of race as socialrace. But it does not preclude the existence of a good correlation between "race" (otherwise conceived) and genetic variability. Root argues that there has been too much interbreeding between 'races', for too long, for the ordinary classifications to capture important genetic heterogeneity and he takes this fact to disqualify race from being a good proxy for genetic variation. Or, he takes the fact that susceptibility to recessive genetic diseases like sickle-cell anemia[49] or Tay-Sachs depends on allelic variations that are localized within particular genetic lineages to disqualify 'race' from being a good proxy for disease susceptibility. But if all we need for a proxy is correlation, why jump ship when there's no causation?

The only plausible reason Root gives is in the ethical arm of his argument: that the risks involved are too high to settle for a *mere* correlation –not that race IS a bad proxy for ancestry.

[49] Root (2003), 1176-1177

But then again, wouldn't we face similar risks if using 'race' to measure social inequalities? Is essentialist thinking about social features any safer than thinking biological traits are essential? Because, as Ellison (2005) points out, this is a clear risk we may run by using "race" in biomedicine as a proxy for social differences.

2.1.4 Using 'Socialrace' as a Proxy for Social Causes of Health Outcomes

So far, the significance of socialrace as a category in medicine has been hypothesized but not explicitly discussed. It is here argued that, although current data is insufficient to "determine the precise relationships that exist between socially defined populations [socioraces] and noteworthy genetic features"[50], phenotypic features linked to disease and treatment response could modulate according to socialrace. In this case, socialrace would be a medically significant category in medicine with special explanatory value.

I will follow the suggestion of Root in the previous section that no study can purport to exhibit a causal relation between socialrace and disease prevalence or response to treatment unless it isolates socialrace as an independent variable by controlling for all possible confounds[51]. The cases that are going to be of interest here are those that control [or try to control] for environmental confounds.

An earlier study on "Racial Differences in the Outcome of Left Ventricular Dysfunction" by Exner, Dries et al[52] (the same investigators performing the ACE study Root (2003) discussed) investigated the possible causes of a higher mortality rate between black and white patients with congestive heart failure. This time the investigators did try to take SES into account by including base-line data on the educational level of patients and the percentage of patients who reported

[50] Foster and Sharp (2002) 845

[51] This is not the only way to go. It can be argued that race is a fundamental cause of health outcomes and not one that can be analyzed as a set of other confounding causes. Public health theorists such as Link and Phelan (1995), (2002) or Jason Beckfield (2004) argue for the existence of root social causes and one might think of race as such a cause. A root cause would be what Cartwright might call a capacity that is always exercised, like say the capacity of massive bodies to attract each other. I.e. a root cause contributes always to an outcome, and in a constant contribution without needing any triggers, and also, it often trumps many other intermediate causes.

[52] Dries, Exner et al (1999)

having experienced "major financial distress" in the year prior to enrollment[53]. More black patients reported financial difficulties and had a lower educational level than white patients, but even after adjustments the black patients remained at higher risk for death from all causes [except arrhythmia which dropped out]. As a result, the researchers "hypothesized that racial differences in the *natural* history of left ventricular dysfunction might also have a role"[54].

This study does not offer direct evidence for a distinct impact of socialrace on mortality rates, but neither does it preclude it. The criteria used for educational level included two categories of more or less than 12 years of education and less than 8 years of education, which seem fairly thorough. However, the same cannot be said of the study's criterion of financial status. Arguably, having experienced some "major financial distress" within the year prior to the study is no good measure of prolonged financial difficulties. What might pass as a major financial distress in the eyes of a well-off white might be everyday reality for a working-class black. Further, it is not only major stress that might influence one's health profile but also smaller strains sustained over long periods of time. Unsurprisingly, univariate analysis associated a lower educational level and not financial distress with an increased risk of death.

This case-study is one where "natural" causes were posited for higher mortalities among blacks without all possible "social" causes having been exhausted. As shown in the previous section, making this causal hypothesis is problematic if biological race is what stratifies the results. What if we take socialrace to classify mortality rates?

There is still no reason to postulate racial differences in the "natural history" of the disease. These results might be just saying that the socialrace of the patient behaves as a *proxy* for other socio-economic environmental influences on the patient's health profile[55]. Even more interestingly, "socialrace" *itself* might be a real, social-environmental causal factor, distinct from SES or other environmental factors that acts independently of any "natural" racial difference to shape health.

---

[53] Dries, Exner et al (1999) 612-613
[54] Dries, Exner et al (1999) 609
[55]This is a view that Michael Root endorses when he talks of race (socialrace) as being a "reasonable" category in medicine, if it's taken to be a proxy of socioeconomic factors.

A possible causal story as to how socialrace might affect health could be found if we accept the relatively plausible claim that one's psychological well-being affects one's health. A recent study by psychiatrist Giovanni Fava and psychologist Chiara Ruini describes the characteristics of a novel psychotherapeutic strategy: well-being therapy[56]. The therapy itself is not of interest here. What is of interest is that this is a therapy for affective (mood and anxiety) disorders i.e. for stress disorders that are known to correlate with high incidences of heart disease, high blood pressure and many more disease phenotypes.

Well-being therapy works via specific psychotherapeutic techniques to enhace the patients' levels of psychological well-being along six identified dimensions: autonomy, personal growth, environmental mastery, purpose in life, positive relations with others and self-acceptance[57]. The therapist "scores" a patient on such levels along provided guidelines for what would count for impaired and optimal levels for these dimensions. For example, if the patient has an impaired level of environmental mastery then he "has or feels difficulties in managing everyday affairs; feels unable to change or improve surrounding context; is unaware of surrounding opportunities; lacks sense of control over external world"[58].

Now, it might be reasonable to assume that well-being could be strongly correlated with one's socialrace. Since socialrace picks out a group that has been historically *taken* to be a racialist race and the non-white socialraces (especially the black socialrace) have hence been and still are the target of racial discrimination and victim to social, political and economic abuse, it might be reasonable to expect that individuals who take themselves and are taken by society to belong to non-white socialraces could exhibit more depleted levels of well-being than whites, have higher frequencies of stress disorders and higher rates of congestive heart disease; and all this partly *because* they belong to a non-dominant socialrace.

However convincing this causal scenario might be, it serves to illustrate the possibility of the existence of an indirect causal impact of socialrace on health. However, in order to even *begin* to contemplate any such status for the category of socialrace, socialrace should first be isolated

[56] Fava and Ruini (2003)
[57] Fava and Ruini (2003) 51
[58] Fava and Ruini (2003) 51

from SES and environmental factors [given that it is already rejected as an adequate proxy for genotypic differences].

The biomedical literature contains other studies which take, or purport to take, environmental factors into account when interpreting racially stratified disease phenotypes. Brancati and colleagues (1996) attempt to "identify factors associated with diabetes mellitus and to determine whether racial differences in these factors, especially socioeconomic status, explain the high prevalence of diabetes among African-Americans."[59] The study indicates that even though SES is strongly correlated with race and is a robust predictor of access and quality of health care, racial or ethnic differences in disease prevalence persist. After controlling for "age, SES, overweight and central adiposity"[60] blacks remained over twice as likely as whites to have diabetes. The excess prevalence of diabetes in African-Americans was, further, found to be greatest in individuals of low SES and least among individuals of high SES. The authors themselves concluded the following:

> The most straightforward interpretation of the observation that African-American race has a strong, independent association with diabetes mellitus is that African-Americans are more *susceptible* to diabetes mellitus compared with their white counterparts. (…) Evidence that diabetes was uncommon among African-Americans prior to 1940 and remains uncommon today among rural West Africans does not refute this notion. These lower prevalences may simply reflect the permissive role of environmental factors –such as modifications in diet and physical activity that accommodate change from an agrarian to an industrial society –in the *expression* of inherited predispositions. Alternatively, race may constitute so strong and complex a marker of behavior and environment in the United States that our adjustment for socioeconomic status, education, and cigarette smoking was unable to account fully for *non-inheritable* differences between African-Americans and whites[61]. [Emphasis added]

This excerpt illustrates how the investigators first consider biological race as a proxy for genetic predisposition for diabetes, how they postulate the role of environmental factors [other than socialrace] as enabling the expression of such unique *genetic* characters, and how they finally consider the possibility that socialrace might be a proxy for environmental causes of higher disease prevalence. This hierarchical ranking of possible interpretations reflects how

[59] Brancati et al (1996) 67
[60] Brancati et al (1996) 67
[61] Brancati et al (1996) 72

unlikely it is in the eyes of these epidemiologists that socialrace is more complex a causal cluster than SES-plus-cigarrette-smoking-plus-education [and how it is unthinkable that socialrace might affect health by independently correlating with individuals' overall well-being].

It seems that, at this point, empirical data is unable to point either way as to the causal role of socialrace as an independent enviromental factor affecting individuals' health profiles. Nevertheless, there is no evidence precluding such a function for socialrace. Socialrace is not an ad hoc notion introduced to bar the possibility of a racially specific genotypic basis for differential health profiles. It is scientific evidence that points towards the inappropriateness of socialraces as proxies for genetic variability. Socialrace just happens to describe a characteristic of our social environment that is strongly correlated with other environmental factors that influence health and could be either reducible to such factors or just one of them.

Distinguishing between "geographical ancestry", which "biorace" might be a proxy for, and a social status like "socialrace" allows scientists to identify two medically significant dimensions along which disease genotypes and phenotypes might vary. Further, the fact that the concept associated with racialist races, *socialrace*, is from the start understood to be social, prevents any interpretation of particular phenotypic characters correlated with socialrace as having an essential, genotypic basis.

Conclusions to Section 2.1 on Root

"Natural selection" of specific gene mutations is one of the ways in which the natural environment influences the genotype and by extension the phenotype of biological organisms. No one, layman or scientist contests the importance of the geographical region of our species' origin in shaping our genotype. Our ancestral environment's particular climate, its flora and fauna, the indigenous germs and viruses all have, and are seen to have, a significant impact on our current biological and health profiles.

However, human environments consist of more than these physical entities. Human environments include social, political and economic structures to which individuals stand in

particular relationships. The "socially constructed" nature of these entities does not preclude them from having a concrete, measurable impact on our lifestyle and consequently on our health. Even if social "strains" operate on the phenotypic rather than the genotypic level their force is no less real. This rings especially true for urban environments. Given the largely artificial and contained "natural" environments of western cities, made up of shaded homes with heating and air-conditioning, early age inoculations and fierce beasts kept in the zoo, in which "natural" strains are rarely experienced, any social pressures felt become all the more "real" and definitive.

This discussion has attempted to identify and distinguish between natural and social forces at play in the shaping of 'racial' health profiles. If we agree with Kitcher (1999) that race is partly socially constructed by virtue of depending on social mechanisms for its endurance [specific patterns of intraracial mating], then we can profess that variability between racial health profiles is also partly socially constructed by virtue of depending on a particular skewed social distribution of the environmental causes of a disease. This does not mean that variability between the health profiles of blacks and whites does not exist. It just means that the disparity between blacks' and whites' health can be significantly eliminated by equalizing the environmental factors that affect it.

Instead of just looking for more nuanced genetic causes for disease, causes that may be cured by some expensive new drug or procedure that functions on the level of genotypic characteristics and disease, scientists should point to and celebrate environmental causal factors affecting health; not as confounds in their search for drugs, but as handles for health policy, as distinct dimensions along which treatment *can* proceed.

Still, I have been arguing that Root's understanding is not the only useful way to understand "race" in the context of biomedical practice. I have in fact argued that we cannot understand race as not genetically important without a biorace notion of race. And we would not want to dismiss the possibility of their being some content in a biorace notion without first examining the hypothesis in biomedical and genetics research.

Indeed, our scientific research can be self-fulfilling, reinforcing its own categories, so choosing which ones to use comes with a certain degree of ethical responsibility on our part. However the choice seems already made in the case of race, at least in epidemiology. And it is a choice that geneticists are having to deal with in their research and –I think- take unjustifiably extra slack for so doing. If we see what is going on not as a reification of an illusory "race" but as a founding so that what bits of it are useful to us for what concerns we have may be used efficiently and with maximum logical strength, then such worries might be mitigated and cooperation made visible as necessary for learning. And if we go with the data of Abdallah Daar and Peter Singer, we might find some ethical grounds for understanding "race" as biorace.

2.2 Abdallah Daar and Peter Singer on the Use of 'Race' for Pursuing Population-Specific Pharmacogenetics

Abdallah Daar and Peter Singer are bioethicists based at the Joint Center for Bioethics of the University of Toronto working for the Canadian Program on Genomics and Global Health. Most bioethicists receive their training in science not in philosophy (or ethics) and so one might expect bioethics arguments to be rather flimsy philosophically. But these bioethicists' argument is by far the clearest and most direct argument of those reviewed in this chapter. This view is distinctive enough and in the right way to typify one way of answering the normative question orienting this debate. This is an argument exactly for using 'race' to approximate medically interesting human genetic variation and –once we clarify the understanding of 'race' here as biorace- I think a rather compelling one.

The distinctiveness of this view lies in a distinctive concern: Where Michael Root worried that race might end up being a poor man's substitute for a "genetic transcript" (2003,1180) Daar and Singer worry that it might not! They argue that addressing racial disparities in drug-response should be a profitable and equitable goal in itself not a mere step towards personalized medicine. They urge pharmaceutical companies to invest in understanding local population genotypes to predict individual drug-response without needing expensive,

individual genetic tests and reason that as developing countries' population makes up 85% of the global human population pharmacogenetics can stand to profit by expanding its product markets.

Despite their expertise Daar and Singer seem mostly insensitive to the ethical but also to the methodological dangers of using a socio-politically loaded and conceptually complex concept like 'race' in a biomedical genetics context. This is because Daar and Singer focus their discussion using a biorace concept of race and as such underestimate the salience of sociorace notions for both ethical but also for methodological and epistemological issues to do with variation recorded under these categories. I take the case of BiDil, the first drug to be approved by the FDA as race-specific, to demonstrate this last point.

Daar and Singer's account would be more accurate were the notion of race under consideration here specified as a biorace notion, and it would be more complete were a sociorace understanding of race factored into some of the normative claims they make. Still this discussion validates the relevance of biorace notions in this discussion and further indicates the full scale to which biorace race notions are usable -and used: cutting through epistemological, biomedical, ethical, business and political questions, such notions seem to align with a host of different interests in the genetic variation of human populations.

2.2.1 The Targets of 'Population'-Specific Pharmacogenetics Research

Daar and Singer (2005) tell us how to organize pharmacogenomics to best align with the interests of all of scientists, patients, stakeholders in the pharmaceutical industry and in the developing nations' governments. This is indeed an ambitious endeavor –but also a very interesting one! Their central thesis is this: An all-around beneficial use of pharmacogenetics relies on studying population as opposed to individual genetic differences.

> In this article, we make two related arguments: first, that pharmacogenetics has significant relevance to the health of people in developing countries; and second, that for this benefit to be realized, we need to take into account not just differences between the genotypes of individuals, important as they are, but the differences in genotypes between different population groups (241).

These claims respond to a worry, which Daar and Singer choose to capture by the metaphor of a house:

> The completion of a good quality draft of the sequence of the euchromatic portion of the human genome [See Appendix 1 section 3 for more detail] was accompanied by a commentary in Nature in which the future of genomics was compared to a house. The question we ask here is: who will live in that house? Is it only the 700 million or so people in the United States and Western Europe, or will the rest of the 6 billion people, who live mainly in the developing world, also be able to find room there? (241).

This is a telling metaphor, given that the right to shelter is acknowledged as a basic human one. It sets a tone for the discussion, and points to certain ethics assumptions the authors make. It is understood as one of our ethical duties to cater to as many people as possible with this science, especially those in greater need –an aim that few, though some, could find grounds to object to[62]. Despite this deontological spur Daar and Singer's argument is utility-oriented. We should cater to these groups, they say, because, as a matter a fact we would all profit from it: science would learn more, patients would get better drugs and the pharmaceutical industry would make more money.

2.2.1.1 'Populations' or 'Races'? Bioraces

Importantly for my thesis, the word 'race' does not appear in Daar and Singer's first statement of their theses; the whole first page of their article talks about "populations" and "population groups". These they use to refer to the ideal targets of pharmacogenetics research. They later will make direct reference to race and ethnicity and endorse using the categories as one available way to approximate these "population groups". Still this indicates that Daar and Singer take themselves to be arguing about a concept that is assumed to be indigenous to contexts of genetics and genomics: 'populations'. This indicates that when they talk about 'race' in this context Daar and Singer understand 'race' as biorace, not sociorace.

> The switch to the first mention of "race" in the article comes as follows:
>
> In January 2003, the FDA called for greater scrutiny of data from subpopulations, asking drug testers to use the racial categories that have been specified by the Census Bureau, to ensure consistency when evaluating potential differences in responses to drugs (242).

[62] A common rationale is that people should earn their keep.

The authors refer to this FDA directive as evidence that legislative pressures support their recommendations for examining group differences in drug response instead of individual ones[63]. It is telling that these authors' first mention of "race" comes in reference to the FDA as if to notarize the receipt of the categories from some "external" higher-level directive, but also and at the same time push for giving them a proper place in genetics.

Indeed the authors proceed to explicitly endorse and even argue for the salience of racial categories for biomedical genomics research. In a section titled "Ancestry and phenotypic differences" they explain:

> Studies in population genetics have revealed a great deal of genetic variation within racial or ethnic subpopulations, but also substantial variation between the five main racial groups which are based on continental ancestry[64] (242).

Racial groups are here referred to as "based on continental ancestry"; a claim that flies in the face of what Michael Root has been arguing. Namely, that there is no basis for thinking of races as biologically meaningful categories. Daar and Singer go on. Following Risch et al (2002) they cite three ways in which the existence of genetic variation between races has been demonstrated:

> [F]irst, ancestral tree diagrams carried out using population genetic data from indigenous groups consistently show that Homo sapiens has major branches that correspond to the five main groups. Second, clusters that have recently been inferred from multi-locus genetic data and other studies coincide closely with groups that are defined by self-identified race or continental ancestry. Third, low frequency alleles are more likely to be race specific. Race-specific variants are particularly common among Africans, who have greater genetic variability than other racial groups but more low-frequency alleles. For *observed phenotypic differences*, self-identified race and continental ancestry often have relatively high predictive power compared to self-identified ethnicity. It is therefore likely that racial or ethnic categories will continue to be useful as long as such categorization 'explains' variation that is left unexplained by other factors (242). [Emphases added]

Whether this evidence is good or bad is irrelevant here. What matters is what the evidence purports to speak to: These three types of evidence are evidence that a biorace notion of race has use for genetics. The evidence is furnished using techniques peculiar to the practice of

[63] It is no accident that the groups hoped to "ensure consistency" are the racial categories specified by the census bureau. This is how ordinary race concepts are found in a domain of specialized biomedical scientific use: they are already available in demography.

[64] What these racial groups are is left unspecified throughout the article but following continental distinctions we might infer Daar and Singer mean: African, European, East Asian, Oceania and the Pacific Islands and American Indian.

population genetics and genomics such as 1) drawing up ancestral tree diagrams using population genetic data and examining where major branching events of these diagrams occur, 2) inferring clusters of genetic structure using multi-locus genetic data and seeing how these correspond to self-identified race or continental ancestry (–here the authors refer to Rosenberg (2002) which is examined in Chapter 3 Section 3.3.2.1) and 3) estimating how low-frequency alleles are distributed across different races.

Further in contrast to Root it is claimed, citing Risch (2002), that both the categories of 'self-identified race' and 'continental ancestry' are better predictors of phenotypic differences than 'self-identified ethnicity'.

2.2.1.2 Why Use 'Race' in Pharmacogenetics?

Why should we assume any of this variation is interesting for biomedical genetics and pharmacogenetics?

The authors refer us to 'Box 1' 'Drug response variation among individuals and populations' for more evidence for this claim.

> During the past 50 years of pharmacogenetics research, we have learnt that variation between individuals that is influenced by genes and other factors is relevant to the efficacy of all drugs. We now know that metabolic enzymes are affected not only by SNPs (of which the human genome contains more than 10 million), but also by other genomic variation, such as gene duplications and deletions, mutations in regulatory genes, and probably by recently-described large-scale copy number variations. Increasing numbers of relevant polymorphisms are being discovered. Most relevant to our discussion, we also know that the frequencies and distributions of harmful and protective polymorphisms vary greatly between human populations (Box 1, 242).

At the last juncture Daar and Singer make reference to relevant studies and these include information collected using race categories[65]. Without a biorace understanding of race, this move from taking results obtained using 'race' to warrant the medical relevance of 'population-group differences' would seem like a logical fallacy.

When Daar and Singer mention genetic influences on metabolic enzymes they do not disbar the influence of environmental factors on these enzymes. And when they say that harmful

[65] See Wilson and Goldstein (2001) for example or a study of Exner et al (2001) on heart failure that Root (2003) dismisses as below par.

and protective polymorphisms are distributed in significantly different ways across 'human populations' and that polymorphisms –as well as other genomic variation such as gene replication and deletions- are important for mutations in regulatory genes[66] and possibly large-scale copy number variations[67] (242) they do not say that environment means zilch. What they argue for is the study of what genetic differences correspond to racial classes, not the dismissal of other influences on health outcomes as uninteresting or irrelevant.

When these authors conclude that racial and ethnic labels can likely "'explain'" variation left unexplained by other factors it becomes clear that Daar and Singer do understand 'race' as a concept that could explain such differences: Even if by putting 'explain' in quotes, they withhold commitment to taking 'race' as a cause it is not just a social status marker like 'marital-status' that 'race' is picking out in this case. For Daar and Singer to think race can be taken to explain (or explain in 'quotes') genetic variation across the so-named groups they must have a notion of the category as biologically interesting and meaningful, which is to say their operative race notion is a biorace notion. Further evidence of this is found in how Daar and Singer fail to imagine and so to anticipate the social, political and ethical problems peculiar to mobilizing these particular categories in a global context.

2.2.1.3 Broader Benefits of 'Population'-Specific Pharmacogenetics

Daar and Singer claim:

0. There are medically important genetic differences between human populations that are roughly approximated by the five 'races'.

And that

1. If we study these differences in genetics and in medicine we can design better drugs for more patients.

[66] Regulatory genes are genes that regulate or define the activity of other genes especially genes involved in protein coding.

[67] The number of copies of a gene people usually have is two, one from each parent, but as a result of mutations such as insertions and deletions some people have three or more copies of a gene and less frequently people may have no copy of a gene. This type of genetic variation is called copy number variation.

Daar and Singer agree with Root that ancient geographical ancestry would well approximate medically interesting differences in human genetic structure; but they also think that ordinary race/ethnicity classifications can approximate these differences, warranting the use of a biorace notion for exploring the biomedical import of genetic variations recorded under race/ethnicity classes.

However the argument of Singer and Daar is broader. It doesn't only answer the epistemological/metaphysical question, of whether there is possibly something for medical genetics to know when looking at "race"-specific genetic structure (where "race" is understood as a biorace). It also discusses why and how this research might practically be pursued. And it views such research as aligned with our ethical duty to cater to a diverse global population, and as in the interest of the pharmaceutical industry to invest in. So, Daar and Singer also argue that

2. Population-specific pharmacogenetics is financially viable.

3. Population-specific pharmacogenetics is ethically commendable.

Were these theses true Daar and Singer's argument for pursuing race-specific genetics and genomics would be strong and one whose strengths would play nicely against Root's argument that there are non-negligible social and political risks that race-specific (medical) genetics research runs for particular socialraces, as well as that argument's lack of attention to matters more practical, like how a use of 'ancestry' -instead of 'race'- might be incentivized and incorporated into current practice. More importantly, however, how Daar and Singer's argument is weak confirms the worry I want to raise, which is that working with only a biorace or only a sociorace notion of race in dealing with broader normative questions leads to trouble.

I will discuss the first of these last two theses, (2) here. The ethics thesis must be evaluated from a perspective that is broader than this dissertation covers [68]. The weakness of Daar and Singer's argument on thesis (2) is enough to make visible the dangers of relying on a biorace concept to the exception of sociorace.

[68] cf. Reiss and Kitcher (2008) for an argument for why what Kitcher calls "well-ordered" science demands that the pharmaceutical industry and biomedical science more generally should target the treatment of common, neglected diseases affecting the populations of the developing world.

2.2.1.4 Financial Incentives for Population-Specific Pharmacogenomics

How might a pharmaceutical company profit from noticing genetic structure at the level of racial populations?

Daar and Singer isolate three ways: a. defining new drug targets, b. what they call "prospective efficacy pharmacogenomics" and c. drug "resuscitation" –or what we may call "retrospective efficacy pharmacogenomics".

First, noticing population-specific genetic structure can be useful in finding new targets for drugs, they say. Pharmacogenetics is reported as adding to its history of fifty years advances that enable pharmacogenetics and pharmacogenomics to join forces, "identify the genetic causes of diseases and search for new drug targets" (241).

> Today, several major pharmaceutical companies have teams that focus their research on the intersection between genetics, genomics and drug development, and some are already beginning to take genomic variation into account in their drug development pipelines. Although the idea of focusing clinical trials on subgroups of individuals is not new – stratification by disease subtype has always been a goal of medical research- the use of genetics in this context is new (241).

Secondly, Daar and Singer say that pharmacogenetic and pharmacogenomic controls at the level of population structure can save the drug industry a lot of money. How? Their suggestions would fall under three strategies. First, there is what they call "prospective efficacy pharmacogenomics" or helping fit drugs to target populations better, so that adverse reactions are avoided. This would help lower any losses caused by having to withdraw a drug, running a costly legal trial to defend a drug, or even losing a market share because of bad publicity associated with a drug's posited adverse effects. Secondly, pharmacogenetics can feature in the post-marketing surveillance of a drug. For example some sub-Saharan African populations have a genetic polymorphism in the ABCB1 gene that encodes the multidrug transporter P-glycoprotein which makes them less likely to benefit from antiretroviral therapy; this finding could save lives and money by not wasting drugs on populations who don't benefit from them and encourage the development of new drugs for these populations (245).

Last, what one might call “retrospective” efficacy pharmacogenomics is finding target populations for already available drugs or resuscitating already FDA-rejected drugs and developing these for newly defined target populations.

To motivate their theses Daar and Singer discuss the case of how the heart-failure drug BiDil made it to the U.S. market.

> … a few years ago, the FDA rejected a fixed-dose combination of isosorbide dinitrate and hydralazine (now known as BiDil, NitroMed) because its efficacy in treating heart failure could not be demonstrated statistically in a clinical trial in the general population. When it was tested exclusively in 1,050 self-identified African-American patients who had experienced heart failure, the results of this double blind, randomized clinical trial were so impressive that in July 2004 the trial (which was endorsed by the Association of Black Cardiologists) had to be stopped for ethical reasons; there was a significantly higher mortality rate in the placebo group than in the group given BiDil. BiDil is now expected to be approved by the FDA in early 2005, as the first ever ‘race-specific’ therapy (242).

Daar and Singer read these FDA decisions as supporting prospective and retrospective efficacy pharmacogenetics along racial population subgroups. Daar and Singer acknowledge that obtaining phenotypic data as related to adverse effects is difficult. But they say that drug companies are already working with the FDA to develop data-mining tools to help them sort through clinical trial data and link it to information from DNA samples (242). For example, the drug company Pfizer is reported to be (at the time of writing) studying hypertension-related genes in African Americans and diabetes-related genes in both Asian Indians and Native Americans that may explain the increased incidence of hypertension in these populations; and the company AstraZeneca is reported as looking for population differences in drug response in clinical trials (242). Daar and Singer say that though these studies are now part of “personalized” medicine programs, in the long run these studies are likely to be more relevant to people in the developing countries (242).

One might accept that in theory Daar and Singer’s suggestions seem reasonable. But do they really have good evidence where ‘populations’ coincide with socioraces? How might marketing a race-specific drug, indeed, not be profitable?

2.2.1.5 The ‘Race’ Liability: The Case of BiDil

This is perhaps a low blow, but a blow nonetheless to Daar and Singer's argument. What actually happened to the market share of BiDil's production company, NitroMed, questions the validity of Daar and Singer's argument. Race-specific drugs need not be profitable for the pharmaceutical industry: in fact they may be especially risky for drug companies to develop because of their "racial" labels.

My analysis does not challenge the biomedical validity of the efficacy claim of BiDil. This is the matter of great debate but a conclusive answer to the debate is not important for my argument. What is important is the structure of the debate itself, and more specifically, the effects of this debate on the marketability and profitability of this drug.

2.2.1.5.1 A Brief Story for BiDil

So, how did this drug come into being? As Jonathan Kahn tells it, it is a telling story. The first methods patent for the combination of the two generic drugs Hydralazine/ Isosorbide dinatrate (H/I for short), later named BiDil, was granted to Jay Cohn for *a general population*. The patent was obtained in 1987 based on the results of the first *Vasodilator Heart Failure Trial*, V-HeFT I (1986) -the first of a series of clinical trials funded by the Veterans Administration (VA) and headed by Cohn as primary investigator. The drug seemed to work *for a general population* in V-HeFT I where borderline significance was obtained (Cohn et al 1986, 1547), but it failed to get a NDA for a general population by the FDA. Not because the drug was found not to work, but because it worked no better than generics. The statistics presented by the company developing the drug, Medco, did not meet the standards of the FDA. The trial was designed 20 years prior to their NDA petition (1997) and because it was a trial run by the VA and designed as a "test of theory" trial (Kahn quoting Cohn) (Kahn 2005, W5-457), the endpoints were not clearly defined and the drug was deemed bio-equivalent to generics.

Medco's stock crashed but Jay Cohn kept pursuing this drug. With the intellectual rights reverted back to him, Cohn and his colleague Peter Carson retrospectively analyzed the V-HeFT I

data on *just* African-American patients to prove that a significant improvement was recorded in this sub-population.

In 2000, Carson and Cohn filed jointly for an H/I methods patent *for African-Americans*. Cohn re-sold the intellectual rights for H/I to a company called NitroMed and obtained a NDA 'pending trial', in 2001, for a drug specific to African-Americans.

BiDil was indeed the first drug to get FDA approval (June 2005) for a particular racial group. It is a drug for heart failure that was found to be effective in self-identified African-Americans. Patients on BiDil and standard heart failure medication experienced a 43% reduction in death and a 39% decrease in hospitalization for heart failure compared to placebo, as well as an overall decrease of heart failure symptoms.

So the same combination, H/I, was covered under a patent for the general population until 2007 *and* is now covered under a patent for African-American Self-Identified Race/Ethnicity (SIRE) until 2020[69].

2.2.1.5.2 What Works And What Doesn't

So, the drug was shown to work.

What founds the claim that 'BiDil works *for self-identified African-Americans'*?

A-HeFT tested the drug on 1,050 patients of self-identified African American race, only. The conclusion drawn was that the drug works, *for African-Americans.* No cohort trial was done to compare 'white' to 'black' patients' outcomes.

Some researchers take this fact to undermine the logical grounds that justify the racial specificity of the drug. They reason that a race-specific drug claim should be warranted by a trial showing that the drug works better –if not only– for the specified race, than for any other[70]. This

[69] BiDil is a combination of generic drugs in a dosage which cannot be obtained independently. (The combination pill cost 1.80$ and the generic pills cost 25 cents each in 2006, which means that though more of the generic pills would have had to be used per day than BiDil pills in a standard dosage, the difference in price would persist as 6 pills per day of BiDil cost 10.80$ as opposed to 18 pills of generics a day which cost 4.50$).

[70] For example, Jonathan Kahn has pointed out that if the same principle is retrospectively applied and given that in many cases study populations were not ethnically diverse and mostly white, many drugs currently on the market should also be branded race-specific. He further argues that the reason we call this

is surely a more stringent demand than what BiDil satisfied. But this demand is not one that a drug with a specified target needs to satisfy.

Any drug needs to demonstrate its efficacy in its target population, say P. To do that, in controlled clinical trials we study what effects another substance, a placebo, might have on the same target. But we don't study what effects the same substance might have in another target group, say P', to establish a claim about a target P. That would normally be beside the point!

So, in showing that 'treatment T works for group "blacks", we would not usually look to show that T works better for "blacks" than for "non-blacks". Further, this would not be needed to logically justify the claim that the drug works for African-Americans. A drug works for African-Americans, even if it works also, or just as well, or even if it works better for another race. The truth-value of the sentence 'A or B' where A is true, is true, no matter what the truth-value of B is.

This remark does not dismiss the importance of debates about whether or not this drug was properly labeled race-specific but it indicates that there is nothing prima facie distinctive about this efficacy claim. It purports to work on the target population that its experimental population was identified to resemble: If A-HeFT worked then presumably the drug worked, for individuals self-identified as African-Americans.

There is even reason why the drug would work *better* in persons identified as of African-American SIRE. The drug is thought to work by supplementing the heart with nitric oxide, which seems to inhibit the structural growth and remodeling of the heart that is responsible for heart failure. It is posited that 'blacks' have a greater deficiency in nitric oxide than 'whites'. Still as Cohn admits "The reasons for that [the nitric oxide deficiency] are not entirely known, but it seems to occur whenever a study has been done"[71]. The nitric oxide deficiency observed in people of African-American *Self-Identified Race/Ethnicity* (or SIRE) could be a result of

---

drug and not these old ones race-specific is because we take the black 'race' of African-Americans to be more manifestly "other" than the white one and so in need of a label. Kahn presented this argument at the conference on *Race, Pharmaceuticals and Medical Technology* held on 4/7/06 in MIT.

[71] See A-HeFT website section on questions about the trials http://www.aheft.org/questions.asp#1 [Last accessed: 12/06/07]

sociorace or biorace contributions to health. Though specifying a disease mechanism is not necessary for treating the disease.

2.2.1.5.3 Failure

That there is prima facie nothing distinctive about how this efficacy claim was tested clinically nor about how it was theoretically warranted suggests that there is not much prima facie logical justification for the extent to which the efficacy of this drug has been debated in the U.S. I say "prima facie" because a second look should point to something distinctive about this case: the drug's target.

There are problems with the efficacy claim that are independent to the methods used to test its truth or falseness: These are problems to do with the concepts used to articulate it; namely, that of 'self-identified African American race/ethnicity' in the U.S. context.

'Self-identified African Americans' specifies a particular group identified by a particular method. Classification uses the method of self-identification to specify race so it corresponds to one 'self-identified race'. This is not specifying an 'other-identified race' and it is not specifying 'inherited race'. But it is still specifying (some) "race". And concepts of race are notoriously contentious and their scientific legitimacy and relevance has been severely contested in recent years.

To get to the crux of this point, the stock of NitroMed that was producing BiDil as its leading drug suffered severely from the controversy surrounding the efficacy of the drug's claim; which disguises the controversy regarding how the targets of the drug efficacy's claim were selected and defined. As a result NitroMed has stopped actively promoting BiDil[72]. Despite the support of the Association of Black Cardiologists (ABC) controversy was fed by a long tradition of racialist -and anti-racialist- thinking in the United States. Though this was evidently a drug that worked for its target population, and worked well enough to be approved in a rush, this target

[72]Charted data on the stock's price since 2004 show a steady decline after 2006 http://investors.nitromed.com/phoenix.zhtml?c=130535&p=irol-stockChart&control_javalowerindicator2=&control_javalowerindicator1=&control_javachartfunctions=&control_javaapplet= [Last accessed 10/09/2008]

population was picked out without sufficient warrant that its "specialness" cut deeper than the political significance of black U.S. socialrace. No biorace concept was explicitly named.

What is the take-home lesson for Daar and Singer's argument?

Daar and Singer failed to predict that a race-spefic drug, like BiDil, would be subject to controversy. As I have already argued Daar and Singer operate with a biorace race notion through most of their argument. A biorace race notion will not account for the loaded social and political history that the groups in its extension may have. This weakens Singer and Daar's argument that population-specific pharmacogenomics will be profitable for the pharmaceutical industry. It imposes an extra demand on companies marketing race-specific drugs: the need to control for the risk of designing drugs to work along categories that are not uniquely understood, contentious and contested in the target contexts where the drugs are marketed.

Surely, one race-specific drug backfiring is no final verdict on the matter. And it can very reasonably be argued that this failure shouldn't count against Daar and Singer's argument (even if they want to use it so) because this was not a pharmacogenomic drug to begin with. Still, one might envision similar troubles for drugs that get race-specific drug approval labels without a careful justification of why the drug's efficacy claim might be "race"-specific and what that race it is specific to is, exactly.

Conclusion to Section 2.2 on Daar and Singer

Daar and Singer admit that there is work to be done before this vision materializes. Besides "knowledge gaps" to be filled they mention problems with the concept of race.

> There are also conceptual and technical problems that need to be resolved, and the use of population groups –at least as currently conceived in terms of race and other unsatisfactory descriptors that conflate with social constructions– is fraught with ethical and social problems that will need to be addressed with interdisciplinary research. The most satisfactory term for population groups at present is emerging as 'geographical ancestry'; but as data accumulate, we may discover other terms –for communities of common ancestry that are more scientifically accurate and that avoid social constructions completely, making it possible to move forward with less likelihood of controversy. We need to change the paradigm from 'race' to human genome variation (245).

Changing the paradigm from ‘race’ to human genome variation echoes the call of Dunston and Royal (2004) (See Appendix 2 Section A2.1). But they do not appreciate the severity of the problem –if that is what ‘race’ is.

Manifestly Daar and Singer speak of “populations”. But don’t be fooled by words. The word “race” and so a race concept does enter this argument and a ‘biorace’ notion of race explains why this argument is not fallacious. Further, that this conceptual dance around “race” and “populations” in reporting and interpreting data is one analysts and practitioners are doing all too easily and coherently (even if cautiously) indicates they are not systematically acting on an empty concept: hence my calling it something! But this also suggests that changing the paradigm from ‘race’ to ‘human genome variation’ is going to be very hard: Why? Because despite the manifest call to arms, this war is already being fought and lost: biorace is there and arguably there to stay.

Where Daar and Singer say that

> the use of population groups –at least as currently conceived in terms of race and other unsatisfactory descriptors that conflate with social constructions– is fraught with ethical and social problems that will need to be addressed with interdisciplinary research (245),

we might add that what problems surround the concept of race are not just ethical and social; they are also epistemological. These problems are not just about dealing with “constructs” and dealing with these carefully because they might backfire. The problem is accepting that these notions are functioning already in genetics and genomics research as scientific concepts.

Trends in NitroMed’s stock could certainly change but the case of BiDil should not so much be a motive for the pharmaceutical industry to develop race-specific drugs as a warning: Keep clearly in sight the risks that come with ‘race’! And look hard, because the risks are hard to discern when using the same terms to study these categories in different contexts.

Should ‘race’ categories be used to approximate medically interesting human genetic variation?

Daar and Singer’s answer to the normative question posed above would be that though ‘race’ should ultimately not be the term of choice for studying human genetic variation, a biorace

notion should be used as long as it is understood as approximating ancestral continental populations.

And note once more that Michael Root would manifestly agree with Daar and Singer that 'geographical ancestry' is what genetic variation across human groups is based on, though the latter see race as a good proxy for ancestry and then link this variation with differences in drug efficacy. Root argues that races are not biological races and that talking about genetic differences across these groups as medically important is nonsense. Even if there are documented differences in how different racial groups respond to treatment it is too hasty to attribute these differences to genetic causes as opposed to badly controlled environmental ones.

The disagreement between Root and Daar and Singer is based on a different understanding of what races are; for Daar and Singer understand racial divisions as based on continental ancestry, think of race as biorace, and conclude it is a rough proxy for medically interesting genetic variation while Root understands racial divisions as the result of contingent social process, thinks of race as sociorace and concludes it has nothing to do with genetics.

3.3 Ian Hacking on the Use of 'Race' for Biomedical Practice

Hacking explains his ideas on classification in a recent article, "Why race still matters" (Hacking 2005). The piece stimulated philosophy of science discussion on the topic (Kitcher 2007) but, as Hacking's work is also read by non-philosophers in the science studies disciplines, it is a view that stands to be familiar to a broader academic audience. The focus of my analysis is two-fold. 1. Hacking's view on classification and in particular on racial classification and 2. what concepts of race Hacking uses in his discussion.

I first discuss Hacking's view on classification, with an emphasis on the case of race. Hacking claims a. that race is a superficial classification, not cutting up the human population into real kinds, but b. that it can still be useful for biomedical practice. He adds that c. whether for real or superficial classifications, Mill's idea that the characteristics that pick out the classification be uniform (in the main) throughout its members needs a statistical formulation. I

will a. explain his statistical ideas, b. compare with Root's notions of a real social kind and a proxy and c. show how his ideas work in examples with race.

I then analyze Hacking's use of "race" and what his putative answer to my normative question would be. I argue that Hacking would have an answer to that question. And he is also it seems in possession and use of biorace and sociorace notions of race. But these two facts are not logically bound up in his argument regarding how to judge the usefulness of race in medicine. Like Root he discusses the usefulness of using races as proxies but unlike Root's his argument does seem true to its word: it does not seem to rely for its normative force on the content of the particular race notion(s) it is formulated by. Rather, Hacking's normative input is premised on how the variable "race" is used/behaves in a particular context; i.e. it does rely on metaphysics but not about race; rather it relies on assumptions used to formulate and deploy the notions of statistical significance, statistical meaningfulness and statistical usefulness in scientific practice.

This possibility mandates discussion. What sort of concept is this "race"? And what type of *scientia* is this science? How does it get its ontology validated without manifest reference to what things are? This is the topic of the next chapter (DIAGNOSIS) and what found science tries to answer.

2.3.1 Hacking on Mill's Distinction Between Real and Superficial Kinds

Hacking asks:

> Why has there been such a pervasive tendency to apply the category of race and to regard people of different races as essentially different kinds of people? (Hacking 2005,102).

He describes a naturalist answer for why race matters:

> One answer says that the distinction is just there, in the world for all to see. Superficial differences between races do exist in nature, and these are readily recognized (102).

Notice the word 'superficial' creeping into the second sentence above. This is not any kind of naturalist view of race that Hacking is describing: it is a "sensible" one.

> The naturalist agrees at once that the distinctions are less in the nature of things than they once were, thanks to interbreeding among people whose ancestors have come from geographically distinct blocks. (…) The naturalist notes that traditional

> racial distinctions are less and less viable the more children are born to parents whose geographical origins are very different.
>
> Sensible naturalists stop there. The belief that racial differences are anything more than superficial is a repugnant error (102-103).

Hacking presents the view of John Stuart Mill on the issue, whom he calls a sensible naturalist about race –or at least someone "as profeminist and antiracist as can be claimed for a white nineteenth-century man" (103).

> Here in modern terminology, is his [J.S. Mill's] doctrine: (1) Nature makes differences between individuals. These differences are real, not constructed. (2) We classify things according to differences we observe. Classifications are made by people and encoded in social practices, institutions, and language. (3) Some classes are such that their members have little in common except the marks by which we sort them into those classes –call those superficial kinds. (4) Other classes have members with a great many things in common that do not follow from the marks by which we sort them into classes. These are 'real Kinds' (103).

So, according to Hacking, Mill's naturalism posits that people observe differences that Nature makes and they make classifications. Some of these classifications capture superficial kinds and some capture real kinds. What's the difference between the two types of classifications?

Classes whose members differ only with respect to a classifying feature and its consequents are superficial. "White things" says Mill, according to Hacking, is an example of a superficial class: It can include a table, direct sunlight and a cloud. These are things that have nothing in common besides the feature used to sort them (and its consequents), i.e. their color (and what they may share because of sharing a color). Hacking quotes Mill [1849, A System of Logic] on the issue:

> White things are not distinguished by any common properties except whiteness; or if they are, it is only by such as are in some way dependent on, or connected with, whiteness (103).

Members of the classes we should call "real", on the other hand, vary uniformly with respect to "great many" features besides the ones used to pick out the class. One of Mill's examples of a real kind is the class of 'horses', says Hacking. Members of the class 'horse' share many more characteristics than what characteristics we use to distinguish them from, say,

squirrels and, further, the features we use to distinguish them like their ability to neigh or having hooves seem unrelated to each other. In Hacking's words

> [horses] have endless properties in common, over and above whatever marks we use to distinguish them from other animals or other kinds of things. Horses form a real Kind, but the class of white things is a superficial kind (103).

OK, so this seems like a sensible observation: some of the classes into which we divide things contain objects that match only with respect to more or less the feature(s) used to divide them and some contain objects that match on more than what feature(s) we use to sort them. This seems like a sensible distinction. But Mill calls one type of class "superficial" and the other one "real". Are these good terms for the distinction? Or, which is to say the same, is this a good definition for what would be real and superficial kinds?

2.3.1.1 A Real Class Need not be Useful and A Superficial Class Might

I already think there is trouble for these two definitions (and Hacking, inadvertently, will reinforce my point). Why?

The problem for Mill's (and Hacking's) definition comes up I think because these definitions are derived focusing on the products of a classificatory process without attention to the process or why it might take place. We are led to agree with Mill's definitions by comparing kinds that are already sorted: told to consider and compare two sets of things (white things, and horses) with no mention as to how these groups get to be sorted under these labels. This assumes that we have good tools for identifying members of these kinds and their features and ways of counting different features; etc. But in fact there is no saying whether this distinction between "superficial" and "real" kinds is one we may entertain in the abstract and whether this is indeed a way of sorting things that will securely give us "superficial" or "real" kinds. Which is to say that Mill doesn't argue to the effect that the distinction between superficial and real kinds is in general well-defined.

One may see problems creeping up already in the case of classifying horses and white things. The trademark of reality according to Hacking's reading of Mill is the "great", or

"endless" number of features distinguishing members of these kinds on top of the features we use to distinguish them. But distinctions are made relative to a background or in relation to a reference class. Say that, in the example I mentioned above we were to compare "horses" with "mules" instead of with "squirrels". If endlessly many more features unify members of the class "horse" than those used to distinguish them from squirrels, could the same be said when comparing horses to mules? What distinguishing features distinguish horses from mules and come on top of what features we use to distinguish the two are arguably fewer than the "endlessly more" features we supposed distinguished horses from squirrels. Are they still enough to distinguish horses as a real kind? Or should we change stories relative to a reference class and just say that horses and mules are superficial kinds of horse-and-mule-like animals?

And how important is the *number* of features used to distinguish members of a class (I will refer to these as sorting traits) relative to the *type* of feature used? What if we were sorting kinds using only one property, as in the case of sorting by white 'color', but this happened to be a property that was remarkable or distinctive of a specific class of things, like 'having blonde highlights'. Admittedly the things collected under those with blonde highlights would be human and so presumably would end up sharing many more properties than what we used to sort them and ones that are of no relation to blonde highlights. But would the kind of 'blonde-highlighted hairy things' be a real one? What if we compared it to the kind 'human'? Wouldn't it be less real than that?

This definition of Mill's of superficiality and reality with respect to classifications primes a particular understanding of what is important in a process of distinguishing between things[73]. This can be problematic because calling this a distinction between what is "superficial" and "real" is loaded. When we call something superficial as opposed to real we could be calling it unimportant as opposed to important; not really-there as opposed to really-there; contingent as opposed to necessary, etc. (Hacking will later use the terms 'significant', 'meaningful' and 'useful' to –I argue- qualify 'reality'.) But I think you will agree that there are some very nice

[73] The idea of connectivity as a mark of an interesting or real classification is widespread; from Hugh Mellor through those who identify events by their sets of causes and effects.

things about the class of “white things” and about the possibility of classifying things as “white” even though this may be a “superficial” kind.

Let me explain. Whether a system of classification is good or useful could be judged on at least the following points: (1) the tightness of fit between feature(s) used for sorting and any one thing sorted, (2) the tightness of fit between members of a class thus sorted, (3) the tightness of fit between the method used for sorting and the features used for sorting, and (4) the tightness of fit between the method used for sorting and the interest or principle for which or by which a sorting method is taken up[74].

Imagine a filing cabinet: (1) you want the labels in your filing system to refer well or clearly to what is inside a file, (2) you want the things inside a file to resemble each other in the way labeled –and perhaps in more ways than the one labeled- and you want things in files with the same label in different filing cabinets to resemble things in your files with the same label [–do you?], (3) you want the labels to label something you are interested in, (4) you want the system to be usable and useful: you want the labels you put on the files to enable you to put the contents of the file to the intended uses.

We may think of these four types of feature as follows, in general, for any classification: (1) as the tightness of fit between a label and a thing, (2) the tightness of fit across things having the same label, (3) the tightness of fit between a label and our tools for picking up labels, (4) the tightness of fit between this label and the reason we wanted or needed to label things in the first place.

A kind that Mill according to Hacking calls superficial, like “white”, differs he says from a kind he calls real, like “horse”, on point (2). But what about the other points? Might “white” trump “horse” on any of these points?

I think it very plausibly might.

For example, how precisely can we define “white” such that things may come to fit the kind? Quite precisely I think: we can pick particular white hues, and use specialized instruments

[74] I do realize that interests and principles are taken to be different things, but they function in the same way in this set up as meta-principles for sorting sorting rules. That is why I include them both under (4).

to look at just a thing's whiteness under just the conditions specified. This seems harder to do for "horse". (And I think notoriously so, because people end up talking about cluster concepts for all the more complex kinds.) Which ones of the "endlessly many" features horses share should we pick? Which cluster? How well or not would things come to fit the label? –Wouldn't we risk getting one or two mules in there? So on point (1) above we may say that we have a better chance of precisely identifying a "white" individual than a "horse". [I am assuming that we cannot easily go about identifying "essences" in the case of horses: so I am assuming there is no say genetic or other test that might in a short script tell us whether x is a 'horse' or not.]

Hacking following Mill makes a big deal out of the fact that no matter how well we may sort individuals by "white", and no matter how sure we are that all these things fit our "white", these things are not necessarily similar to each other in any other respect. But then again, that may be of great interest and value too! Let us think of some such reasons. Discount a prior interest in basic defining features of the kind, such as looking at things that say, reflect light in such and such a frequency under such and such conditions, the "such and such" specifying a class of white things as opposed to horses. Say I throw a party.

I have many good reasons to pick "white" as opposed to "horse" as the party theme. Why? It is a more flexible one. One can comfortably dress up in different types of clothes, of different material, tailoring or style, and fit my party's theme. If the theme was "horse", well, unless we interpret it freely, my guests would end up looking quite foolish and I don't think they would like that. In short, colors may tag all sorts of things and this is something very useful about them, and something we cannot say of classes like species. This is a point I described under number (4) above. But I gave an example about parties!

Take then an example from the practice of biochemistry: the example of dye-terminator molecules used in dideoxy sequencing of DNA (discussed in Appendix 1, Section A1.2). Frederick Sanger was awarded what was his second Nobel prize in chemistry for inventing a technique at the heart of which lies a simple color-coding scheme. To remind you, the technique uses terminator dideoxy nucleotides that are chemically manipulated so that no more bases can

bind to them (and so DNA polymerase stops filling in a template strand at some point) but also dyed with one of four colored dyes corresponding to whether the base is adenine, thymine, cytocine or guanine. Complementary strands to the same DNA template are first sorted by length via capillary electrophoresis, dyes corresponding to each base are excited by a special laser and the frequency of light they emit is detected by a photocell and recorded in an electropherogram. Reading the order of bases in a DNA strand is effected by means of reading the order of colors that come up in this electropherogram.

You must admit that this is another cool fact about a color: It can "dye" a molecule –as well as a cloud or a table. And also, it can be detected by a photocell. I am not saying that innovative science cannot be obtained by using horses as proxies for something of interest… [perhaps as proxies for where the pastures are greener? –Enter some game-theoretic technique]. But the special looseness of a kind like "white" is here of extra use –and so its "superficiality" is not such a bad thing after all.

This example demonstrated a point I described under number (4) above: an interest in studying DNA accords well with an interest in figuring out the order of bases in a DNA strand, which fits well with an interest in the order of colors in an electropherogram. But notice too how the importance of tools under point (3) becomes visible in this example: it was after the invention and wider use of lasers that we happened to have the right tools to read out the color "labels" used to tag DNA nucleotides, which made it sensible to choose this kind of label as opposed to another.

Mill defines superficial classes as classes whose properties are only those we use to sort them, and their consequents. But though there may be only one property distinguishing classes, it may be an important one, or a deep one. The fact of the matter is that not all properties are alike. Which brings us back to a talk of "essences". What if there are some unshakeable features, be it for Nature or your interests: features that we want to think of as important? And what if the defining features we use to pick out (superficial) kinds, end up being in such a way "essential", and in some way considered causally responsible or causally necessary for the having of all sorts

of other features; then we may have to end up calling a superficial kind a real one: or real *for us*. We can make lucky 'right' choices or calculated 'wrong' ones. Which suggests that the terms "superficial" and "real", even if taken to label ends of a scale rather than being yes-no matters, may not always be the relevant way to describe objects thus distinguished[75].

This discussion may have sounded like a big diversion but it is not. Though Hacking in word follows Mill, talking about the superficiality of 'race' and how sensible naturalists stop at the surface, in practice Hacking is thinking like me. He has in mind process, not just outcome. So though he focuses on and accepts the importance of (2) explicitly, his conclusions challenge Mill's distinction -and this I think is because he would take some of the points, parsed under (1), (3), (4), quite seriously.

2.3.1.2 Mill's Naturalism Compared to Root's Realism

What would Mill according to Hacking say about races and how is Hacking using Mill's distinctions? First, Mill proposes that races –like sexes- may be real kinds: only empirical science could answer this question.

> If their differences can all be traced to climate and habits [or, he added in later editions, to some one or a few special differences in structure], they are not, in the logician's view, specifically distinct (Hacking quoting Mill, 104).

Hacking explicitly proposes that we use Mill's formulations to think about races. He concludes that racial classifications such as 'black' or 'white' can be seen in a naturalist light as picking out differences that are *there*, created by nature; but concludes, rather summarily, that these are differences corresponding to only superficial classifications.

> Science might have revealed an endless number of differences between the races that are not consequences of the marks by which we distinguish them, namely color and physiognomy. But science has not done so, and almost certainly will not (104).

Note here that Hacking is referring to races as distinguished on the basis of color and physiognomy; this would be a notion that is biological in a basic sense insofar as it fixes its

[75] The importance of criteria (1)-(4) needs further scrutiny. Why a superficial class may fare better along criteria other than (2) could perhaps be attributed to what we may be think of, keeping up with the superficiality metaphor, as a kind's "shininess" or "dullness"; so a shiny superficial class would be one that is good along 1,3 and a dull real class one that is bad along 1, 3. Criterion (4) has to do with relevance.

referent via reference to features of a biological sort. So this is a biorace race notion that Hacking is referring to as the "race" that science did not reveal as real.

So far, Hacking is using Mill's vocabulary to talk about 'race' and he calls this "superficial". Now there is an interesting point to be made using the distinction between a biorace and a sociorace notion of race. Recall Michael Root's definition of what a natural category is and my abstraction and formulation from his sayings of what a real social category could be (Section 2.1.2.1, this chapter):

(RNC) "A naturally occurring category K is real if and only if K makes extrapolations of many discoverable traits possible across all K things" (S633).

(RSC) A social category K is real (if and) only if K makes extrapolations of discernible traits things K ought to have according to subjects and/or norms in the context in question (and K enters into significant though local generalizations).

Juxtapose these with what I think would be formulations of Mill's definition of real as opposed to superficial kinds[76]:

(MRK) A kind K is real if and only if members of K share endlessly more traits than the traits used to sort them into K.

(MSK) A kind K is superficial if and only if members of K share no more than the traits used to sort them into K and consequences of these traits.

I assume we can take the notions –and so the terms– 'category' and 'kind' to be interchangeable here. Note too that Hacking does not use the term "natural" as he doesn't think it is a good one to use. Under Hacking's reading, Mill is not distinguishing between natural and social kinds. Such a distinction would have to do, if we follow Root who follows Elster, with the source of normativity bounding things into reality: so natural kinds are bound by laws of Nature whereas social ones are bound by laws of Man.

[76] I am not here attempting to improve on Mill's definitions. I think that if we specify a set of sorting traits Ts as part of the set of traits T shared by a kind and think about it mathematically, perhaps as dense in T in the case of real kinds, then the phrases "endlessly more and" and "no more than" can be quantified and limits to our epistemic access articulated; further the process of sorting can be further described by looking for relations on this set (as opposed to the set of features of the kind). [This attitude towards naming is different to expecting some exhaustive description.] This is I think in effect what Hacking does later in his paper.

We then get a way, it seems, to be able to talk about both real and superficial, natural or social kinds:

(RNK) A kind K is a real natural kind if and only if members of K share endlessly more discoverable traits than the discoverable traits used to sort them into K.

(SNK) A kind K is a superficial natural kind if and only if members of K share no more than the discoverable traits used to sort them into K and discoverable consequences of these traits.

(RSK) A kind K is a real social kind if and only if members of K share endlessly more traits, prescribed as traits they ought to have according to subjects and/or norms in the context in question, than the traits, prescribed as traits they ought to have according to subjects and/or norms in the context in question, that are used to sort them into K by subjects and/or norms in the context in question.

(SSK) A kind K is a superficial social kind if and only if members of K share no more traits, prescribed as traits they ought to have according to subjects and/or norms in the context in question, than the traits, prescribed as traits they ought to have according to subjects and/or norms in the context in question, that are used to sort them into K by subjects and/or norms in the context in question, and consequences of these traits according to subjects and/or norms in the context in question.

Or to simplify the last two definitions

(RSK) A kind K is a real social kind if and only if members of K share endlessly more prescribed traits than the prescribed traits that are used to sort them into K by subjects and/or norms in the context in question.

(SSK) A kind K is a superficial social kind if and only if members of K share no more prescribed traits than the prescribed traits that are used to sort them into K by subjects and/or norms in the context in question and consequences of these traits in the context in question.

It becomes obvious that there are more permutations possible in the structure of these definitions. Should traits shared by members of a kind besides the identifying ones be of the same

type –whether they are endlessly more than these identifying traits or not? Is the type of kind we're discussing fixed by the type of identifying trait, and is the fact that this type does not exhaust the rest of the traits shared between members of these kinds why we get confused in where to locate a kind's "origin"? Is this why we think, for example, that the kind 'race' cannot be both a (superficial) natural one, and a (real) social one?

Recall that Root ventured to talk about real *social* kinds because he wanted to justify why we could and should be studying race in biomedical research, even though the kind is not a real natural one. Hacking –*pace* Mill- calls race a superficial kind. But Root calls race a real kind – even if he qualifies the reality with a 'social' label.

So, who is right?

Both!

There is a sense (or rather two) in which the two views cohere: if we distinguish between sociorace and biorace understandings of the concept 'race'. Root calls "race" a real social kind and Hacking calls "race" a superficial kind but given previous analysis we can infer that Root is calling race understood as *sociorace* a *real* social kind and that Hacking is calling race understood as *biorace* a *superficial* natural kind. Root would agree with Hacking that biorace is a superficial natural kind and I think that Hacking would agree with Root that sociorace is a real social kind (as becomes apparent in the second section of Hacking's paper).

But all this I say given that we go with a distinction between reality and superficiality as described by Mill. I said that Hacking updates this frame and that he inadvertently wrecks it – though he perhaps gives us some building blocks to build a better frame in its place.

Recall that Hacking's question was a general one: Why does race still matter (in general)? In this paper Hacking is not only thinking about the significance of race for medicine. After the brief rendering of Mill's definitions of real and superficial kinds, the suggestion that whether race was a superficial or a real kind was a matter of empirical science and the declaration that current science has disproved the reality of race, Hacking adds:

> This conclusion, however, does not answer, or aim at answering, the specific question I raised at the outset, of why there is such a pervasive tendency to apply

> the category of race. Maybe Mill thought the answer was obvious. The desire of one racial group to dominate, exploit, or enslave another demands legitimacy in societies that, like modern Europe and America, are committed to versions of egalitarianism. Race sciences were devised to discover a lot of differences between races that do not follow from the marks of color and structure by which we distinguish them. You do not have to treat people equally, if they are sufficiently different (104).

This is the problem: Hacking doesn't want to explain why it is possible to classify by race in this paper but rather why we do it to such a degree (in particular, why we have been doing it for so long). He also states here, foreshadowing his later discussion, what "the answer" (that Mill might have thought obvious) is: discovering lots of differences between the races on top of the marks of color serves to license treating different people differently and serve different interests, while keeping with the egalitarian mandate to treat equals equally.

So there you have it, we traipsed from talking about naturalism and classification to talking about power, in true Foucauldian fashion. But despite this coda at the end of the section ending on page 104, the naturalist in Hacking can't help but poke at some "recent events".

2.3.1.3 Hacking's Updates to Mill's Naturalism

Hacking asks to be pardoned for the diversion and notes that "some recent events force us to clarify the naturalist position on race". And this is I think where he clarifies Mill's position to the point of rarefaction! After an exposition of a seemingly straightforward distinction between real and superficial kinds and a quick verdict that race is a superficial kind, Hacking catches himself.

The general urge is that "We must first update Mill with a little logic" (104). In 1843, when Mill was writing, statistical differences "were only just beginning to loom large on the scientific horizon" (105) so Hacking takes it up on himself to update Mill with some of the logic of statistical differences.

A first goal is to distinguish Mill's understanding of difference as "uniform difference" from what would be that of "statistical difference".

> When he [Mill] wrote about differences between classes, he had in mind properties that serve to distinguish members of one class from another in a uniform way. A

> uniform difference between cows and horses is something that is true in the main of any cow but not true in the main of any horse –digestion by rumination, for example. There are ever so many such differences between horses and cows; hence they are real Kinds. Call them *uniform differences*. There are a great many uniform differences that distinguish horses from other kinds of animals, but almost no uniform differences that distinguish white things from green things, except their color, or Muslims from Christians, except their faith (104).

For any two kinds, uniform differences would be *almost* like what Hardimon earlier referred to as typological differences (Chapter 1; Section 2.1). I say "almost" because the characters according to which the members of two kinds differ, uniformly, are not required to be true of each and every member of the one kind, and false of each and every member of the second kind. Even in the case of defining uniform differences Hacking qualifies his claims by speaking of truth 'in the main'[77]. So, what distinguishes real from superficial kinds says Mill, if we follow Hacking, is that the sorting traits by which we distinguish them are traits of this "uniform" type.

Hacking says however that to understand a modernized naturalist position on race "We need some new concepts: I will use the words 'significant,' 'meaningful,' and 'useful.' All three go with the dread word 'statistical'" (105). He adds that

> Since we are among other things talking about so-called races, namely, geographically and historically identified groups of people, we are talking about populations. And we are talking about some characteristic or property of some but not all members of a population (105).

By "races" we are making reference to "geographically and historically identified groups of people" and so to "populations" and what we are talking about when we are talking about races are characteristics or properties of some members of these populations, he says. These are specific steps in Hacking's logic that allow the notion of a racial difference to become smoothly embedded in a new statistical context; they are non-trivial steps.

Why does Hacking say that the properties we're talking about, in the case of race, do not typify the whole group?

He doesn't give any examples but he has good reason. It is a common thing to say when talking about races that the differences we commonly use to pick out 'race' don't vary uniformly

[77] More formally, given kinds X and Y we can think of the difference between the kinds as a function D(x,y) defined on the Cartesian product of the two kinds XxY, such that for any members x of X and y of Y (or traits x of members of X and traits y of members of Y), D(x,y)=|x-y|. Then if D is a *uniform* difference between X and Y the value of D would *almost* never be zero on XxY.

between racial classes: for example, skin color, the most popular of racial traits, can be the same in members of races we would call different ones. Think of the skin colors of a white person who is tanned and an Asian person who is not, for example: they could very well match. And you can imagine that this can happen with other racial traits as well[78].

Would Hacking have reason to limit his talk to just racial traits and races? That is not clear. It seems that all kinds of classes suffer from exceptions to the rule and if we judge by Hacking's careful reference to difference "in the main" even in the case of real kinds this is not a point lost to Hacking –nor to Mill. Think of for example monster births, which we'd still call members of a specific kind; a horse with 2 tails, is still a *horse* with 2 tails. This is a first indication that one may have reason to consider the notions of statistical difference, which Hacking defines here, as applying to all kinds of kinds.

In any case, variations are real in the case of race, he claims, so Hacking proposes to resolve this problem by abstracting from Mill's talk of uniform differences between classes to talk of statistical or average differences.

2.3.1.3.1 Hacking's Statistical Concepts

Hacking first defines what he calls 'statistical significance':

> 'Significance' was preempted by statistics early in the twentieth century. It is completely entrenched there. Here I use it for any major difference detected by a well-understood statistical analysis. A characteristic is statistically significant if its distribution in one population is significantly different from that in a comparable population (Hacking 2005, 105).

Hacking defines the 'statistical meaningfulness' of a statistically significant characteristic as follows:

> Let us say that a characteristic is statistically meaningful if there is some understanding, in terms of causes, of why the difference is significant (105).

[78] Note that this is different to another claim made often in this literature (e.g. Zack 2002) that skin color varies continuously along racial lines, with no clear cut-off points for calling one 'white' as opposed to 'non-white'. A problem that I think is relevant to all sorts of classifications as it borders on a sorites paradox.

Hacking gives the example of how the association between smoking and lung cancer was initially only significant, statistically, but has since become (statistically) meaningful because we now understand that the association is causal.

> For example, in the early days no one know why smoking was associated with lung cancer, but now we understand that quite well, although not completely. The correlation used to be merely significant, but now it is meaningful (105).

Before we can call a characteristic statistically meaningful we need to have a causal account of why there is a statistically significant difference in the distribution of the characteristic across our study population. 'Having lung cancer' is a statistically meaningful characteristic across smokers and non-smokers because we understand that smoking causes lung cancer. Or similarly, 'smoking' is a statistically meaningful characteristic across lung-cancer sufferers and non-lung cancer sufferers because we understand that smoking causes lung cancer.

Hacking last defines 'statistical usefulness':

> Finally, a characteristic is *statistically useful* if it can be used as an indicator of something of interest in some fairly immediate practical concern. Take an example from another topic nowadays much discussed. A body mass index (BMI) over 31 is a statistically useful indicator of the risk of type 2 diabetes, and is therefore useful in epidemiology and in preventive medicine. (There are much better indicators involving the distribution of mass and muscle in the body, but at present such indicators are expensive to measure, while BMI measurement costs almost nothing) (105).

Hacking doesn't go into a discussion of what it is to be an "indicator" but his definition of a statistically useful characteristic echoes Root's definition of a "proxy" (RBP):

> (RBP) Let *Y* be a variable that is material to an interest *I* but that cannot be directly measured, and *X* a variable that can be directly measured but is not material to *I* but correlates with *Y*. In that case, *X* is a proxy for *Y* if *X* is used instead of *Y* in making a decision about the individual in order to further *I* (Root 2003, 1175-1176).

Hacking's definition is looser than Root's but I think that they are talking about the same thing. Indicators are ways to approximate a variable of interest in an expedient manner. So, one way to define what an "indicator" would be, or when a characteristic might be properly used as such, is to use Root's definition of a proxy. Hacking's definition would thus be rephrased as follows: Something is a *statistically useful* characteristic if it can be used as a *proxy* for something of interest in some fairly immediate concern.

Hidden in Hacking's definitions is the assumption that there are definable populations that can be considered as comparable along some, and the same, characteristic. To make this visible let me rearticulate these definitions as follows:

> Given two comparable populations, say P1 and P2, that share a characteristic, say c, then
>
> (HSSc) c is a *statistically significant* characteristic if and only if the distribution of c in P1 is significantly different from its distribution in P2.
>
> (HSMc) c is a *statistically meaningful* characteristic (relative to P1 and P2) if and only if c is statistically significant and its variation between P1 and P2 is understood in terms of causes.
>
> (HSUc) c is a *statistically useful* characteristic if and only if c can be used as a proxy for a feature material to some fairly immediate concern.

Hacking does not specify whether a statistically useful characteristic c would be statistically significant. From the definition of a proxy we know c will be correlated with a feature material to us, but it is left opaque what that feature is. So BMI might be a statistically useful indicator for the risk of type 2 diabetes, without this implying that BMI is statistically significant characteristic of the populations we are studying.

Manifestly this is how Hacking defines the notions of statistical significance, statistical meaningfulness and statistical usefulness: on *characteristics* of Classes –hence my tagging of these definitions by the lower-case "c" above. But after defining them on characteristics picked out by a classification, Hacking flips over to talking about the Classes having these characteristics in terms of the same notions. He doesn't say how this flipping works but this is what he does: Where c is a statistically significant characteristic relative to the two classes P1/P2, he calls P1/P2 a *statistically significant class* with respect to c –and he likewise talks about statistically meaningful and statistically useful classes, when sometimes, where P2 is the complement of P1 (relative to some assumed population P1 U P2), he represses reference to P2:

> e.g. …*classes* that are statistically significant, meaningful or useful are not thereby real Kinds (105), [Emphasis added]

> …the *class* of people who smoke was known only to be statistically significant with respect to lung cancer (105), [Emphasis added]
>
> …races employed by traditional racists might be statistically significant *classes* (106), [Emphasis added]
>
> …[Murray and Herrnstein] argued that the *class* of African-Americans is a statistically significant *class*- significant with respect to a property they called intelligence, and which they measured with IQ tests (106), [Emphasis added]
>
> … Ashkenazi is a statistically significant and a statistically meaningful *class* with respect to Tay-Sachs disease (107). [Emphasis added]

In fact, when analyzing the medical discourse on race this is the only way he uses statistical significance, meaningfulness and usefulness: to talk about *classes*.

But this is quick!

Reconsider Hacking's quotes:

> …the *class* of people who smoke was known only to be statistically significant with respect to lung cancer (105), [Emphasis added]

For Hacking this means that 'having lung-cancer' is a statistically significant characteristic with respect to population P1= smokers and P2= non-smokers. Or:

> Murray and Herrnstein] argued that the class of African-Americans is a statistically significant *class*- significant with respect to a property they called intelligence, and which they measured with IQ tests (106). [Emphasis added]

For Hacking this means that IQ is a statistically significant characteristic with respect to population P1= African-Americans and P2= whites. Or:

> … Ashkenazi is a statistically significant and a statistically meaningful *class* with respect to Tay-Sachs disease (107). [Emphasis added]

For Hacking this means that Tay-Sachs is a statistically significant and statistically meaningful characteristic with respect to population P1= Ashkenazi Jews and P2= not Ashkenazi Jews.

Hacking's shorthand can be risky, especially his emphasis on only one of the sub-populations with respect to which statistical significance is defined. It takes us from comparing the distribution (meaning/use) of a characteristic shared across classes to comparing classes with respect to that characteristic. To call population P1 statistically significant with respect to c can misleadingly suggest that what distribution c has in P2 may be considered as a baseline or a norm

when, strictly speaking, both classes are statistically significant with respect to each other regarding c. This doesn't mean that there can be no acceptable criteria for choosing what a reference class may be; only that these are decided by considerations external to the formal setup provided by Hacking and that such considerations may vary depending on what background interests guide an analysis.

For example, we may understand why, instead of considering the distribution of the characteristic 'Tay-Sachs' [a hereditary neurological disease caused by an enzyme deficiency] across classes including 'Ashkenazi Jews' and concluding that Tay-Sachs is a statistically significant characteristic across the classes 'Ashkenazi Jews' and 'non-Ashkenazi Jews' [to use a simple parsing of the population space], Hacking flips to talking about the significance of a particular class, 'Ashkenazi Jews', with respect to Tay-Sachs. He does that because the incidence of Tay-Sachs is significantly higher in Ashkenazi Jews than in the other classes and he is interested in classes that risk having a dangerous disease, not those that don't. And this is presumably because we are in general more interested in (avoiding getting) a disease than its absence.

In any case, following Hacking's usage of these terms, and to keep our notions straight we may attribute the following definitions of statistical significance, meaningfulness and use, for any given *Classes* P1 and P2 and characteristic c to Hacking:

> (HSSC) P1 is a *statistically significant* class with respect to characteristic c if and only if the distribution of c in P1 is significantly different from its distribution in some implied comparison population P2.
>
> (HSMC) P1 is a *statistically meaningful* class with respect to characteristic c if and only if c is statistically significant and its variation between P1 and some implied comparison population P2 is understood in terms of causes.

There are two possible formulations for the next definition.

Besides flipping (HSUc) into

> (HSUC) P1 is a *statistically useful* class with respect to characteristic c if and only if *c* can be used as a proxy for a feature of some fairly immediate concern

consider this formulation:

> (HSUC)' P1 is a *statistically useful* class with respect to characteristic c if and only if *P1* can be used as a proxy for c, where c is a statistically significant characteristic of P1 relative to some implied comparison population P2 and of some fairly immediate concern[79].

Hacking did not earlier define the statistical usefulness of a characteristic as contingent on its statistical significance; but the possibility is a live one.

The definition assuming statistical significance is closer to Hacking's usage of the notion of statistical meaningfulness in the case of race. What follows in (HSUC)' from c having a statistically significant different distribution in P1 from P2 is that c is correlated with 'is a member of P1'. This is how we couched the notion of a proxy. So what gets added when we say 'Ashkenazi Jew' is a statistically useful class with respect to Tay Sachs is that Tay Sachs is something we care about.

How would this discussion relate to what I called an apt use of a proxy? I claimed that the question of whether *X* is *aptly* used as a proxy for *Y*, where *Y* is material to interest *I* but hard to measure, we need to consider:

1. The strength of a correlation between *X* and *Y* (which I gather Root assumes happens when there is a causal story relating X and Y) and

2. The importance of the interest *I* for the individual (or other agent) –so that we can decide whether a rough approximation for *Y* by *X* is better than no means of approximating *I* at all.

These two dimensions could also map onto dimensions along which statistical significance and statistical use are decided.

---

[79] The second formulation I owe to Nancy Cartwright.

2.3.1.3.2 What is Updated?

Hacking calls this an “update” of Mill’s naturalism. But what about Mill’s apparatus is updated here?

We’re still talking about three ways to distinguish classes –in terms of their statistical significance, their statistical meaningfulness and their statistical usefuless–one natural question: Is this a sophisticated way to distinguish between what Mill called “real” and “superficial” kinds?

No.

Hacking admits that “Classes that are statistically significant, meaningful or useful are not thereby real Kinds” (105). Why not?

Simply because the notions of statistical significance, statistical meaning and statistical use are defined on classes only via particular characteristics; but there is no telling how important (shall we say significant, meaningful and useful?) these characteristics *really* are. For example, in the case where the characteristic is Body Mass Index (BMI) “There is no reason to believe that there are a great many independent and uniform differences that distinguish obese persons from those whose BMI is in the recommended range of 18 to 25”(105).

But let us think about this comment more carefully. Real kinds could be distinguished from other real or superficial kinds by characteristics that are statistically significant, meaningful and useful. For example, horses may differ from squirrels with respect to ‘having hooves’ significantly, and there could be a causal explanation why ‘having hooves’ differs significantly across these two kinds. The notion of statistical use is defined relative to the situation at hand so, if ‘having hooves’ is a good proxy for being the kind of thing that –say– I can load with my luggage, then ‘having hooves’ might be a statistically useful characteristic for me to measure. So ‘__ is a horse’ could be statistically significant, meaningful and useful with respect to having hooves. Would this type of reasoning be enough to conclude that ‘horses’ are a real class? Not really.

Say I had picked the characteristic 'number of eyes'. This would presumably not be a characteristic that varies significantly between squirrels and horses (though there might be an explanation of why it doesn't vary). It would also most likely not have been much use to know about 'number of eyes' if what I cared about was to get my luggage carried. Which means that, had I sorted things squirrels-and-horses by 'number of eyes', I would not have distinguished horses from squirrels as of particular significance, meaning nor use. 'Horses' and 'squirrels' would end up in the same class. I could conclude that these are not classes that are *real* with respect to number of eyes –but this is not the use of 'reality' that Mill has in mind[80].

Similarly, superficial kinds could but need not be distinguishable from other superficial or real kinds in terms of statistically significant, meaningful or useful traits. 'White' things would likely vary significantly from 'green' things with respect to the frequency of light emitted from their surface, and there is an explanation as to why 'frequency of light emitted' varies significantly between them. Further, if 'frequency of light emitted' is a good proxy for brightness and I'm looking for running gear that stands out in the dark, the feature could be statistically useful. But pick a characteristic like weight, had by all (white or green) things, I presume, and you would most likely find that its variation across white and green things is not significant nor meaningful and so not useful for practical purposes.

What if we only looked at the set of traits that we were supposed to sort kinds by? Presumably this would not include 'number of eyes' were our reference class that of squirrels-and-horses. We need a reminder.

Recall Mill's definition of real as opposed to superficial kinds:

(MRK): A kind K is real if and only if members of K share endlessly more traits than the traits used to sort them into K.

(MSK): A kind K is superficial if and only if members of K share no more than the traits used to sort them into K and consequences of these traits.

[80] Though it might be the promiscuous reality that Dupre (1993) (2002) has in mind.

Say we have a real kind K. Now take Ts to be the set of sorting traits by which we sort members of K into K. Ts is only a subset of the "endlessly more" traits that the real kind members share, call it Tg, as the set of traits specified by some all-knowing, all-grasping sorter, 'God', and say this is the real set of traits these kinds share, Tr. First of all we know that –given we've got our sorting scheme right, Ts is a subset of Tr[81]. We can think of how Tr might be "endlessly more" than Ts mathematically, perhaps by saying that Ts is dense in Tr, though this means that we will never really figure everything out (roughly speaking, between each of the traits in Ts there will be a trait in Tr).

In any case, I gave you this picture of traits in sets, all mixed up, real and sorting ones, so that we may now think better about Hacking's distinctions. Question: What gets in Ts?

One answer, and I think a plausible one, is that the characteristics in Ts ought to be statistically significant, meaningful and useful, as Hacking defines the terms. But we might also expect Ts to include other characteristics that are neither statistically significant, nor statistically meaningful, nor even useful. History would prove me right on that, I think. That is, unless we conceive of 'significance', 'meaning' and 'use' in relative terms. Then perhaps, Ts would indeed only contain features that are (taken to be at each sorting episode) statistically significant, statistically meaningful and statistically useful.

So the key question now is how does Ts relate to Tr? Answer, Hacking's distinctions are used on our way to grasping the full set of REALLY significant, meaningful and useful traits, "really" because God thinks so: the real traits, Tr. Ts tends to Tg in the limit of 'our' notions of significance, meaning and use tending to 'God's'.

And so you see why I suspect that this apparatus is one for thinking about kinds in a rather more promiscuous fashion than Mill originally intended. And why I think Hacking's distinctions are still quite useful.

That Hacking's distinctions don't map onto those he attributes to Mill is not surprising. Look at how he says Mill distinguishes real from superficial kinds: Mill compares classes, sorted

[81] In the worst case the two wouldn't even intersect.

by characteristics. Hacking compares characteristics, defined on classes. Hacking's logic might in a sense pick up from where Mill's leaves off. If Mill looks at classes having specified sets of characteristics as 'the ones we sort classes by' his would be a coarser-grained and rather more phenomenological approach than Hacking's. Hacking is instead picking up one of these features (of interest) defined on some classes (of interest) and asking 'is this feature significant, meaningful and useful as defined on this class?' –let's count/ask/use… And he does so. And then he uses the answer to say whether some class is significant, meaningful and useful with respect to that feature. This is a type of loop, possibly. A loop whereby we can "update" our set of sorting traits, to best approximate… well, what our notions of significance, meaning and use dictate.

I take it that Hacking offered here the notions of statistical significance, statistical meaningfulness and statistical use in hope of cashing out the import of uniform difference across its numerical, causal and practical dimensions. The hope is to get at features that distinguish classes (real or superficial ones) as uniformly as possible; in other words, the hope is to get at what I earlier (section 2.3.1.1) specified as characteristics of a useful classification –irrespective of whether it's one of 'superficial' or 'real' kinds. Hacking's hope is too to get at classifications useful for different interests.

This is a worthwhile effort, especially when the differences between kinds are still unclear –as seems to be the case with racial classifications. But if we decide to look at differences in terms of statistical concepts the distinction between real and superficial kinds becomes one of degree rather than kind: both superficial and real kinds will have characteristics that are statistically significant, meaningful and useful, so much is clear.

Further, Hacking is not entitled to just call races superficial kinds anymore. Unless he can describe superficiality in terms of the new statistical concepts; how many more statistically significant, meaningful and useful traits do real kinds have as opposed to superficial kinds? 'Endlessly more' doesn't ring like a good enough answer anymore. Not unless we accept that there might be endlessly more agents with respect to whose notions of significance, meaning and

use we may need to compute these. Are real classes endlessly more significant, meaningful and useful than superficial ones? And why so?

We could abandon the distinction between superficial and real kinds. The benefits are clear: we have a unified approach in measuring differences and don't have to make a judgment at the end of 'counting' about whether or not what we're measuring is superficial. What seems lost is the space for change that superficiality reserves. Calling a kind superficial means that the traits it captures don't reach the essence, the 'core' of the stuff described –an option that we'd like to keep open when talking of racial classifications… [and may be true of all classifications?]

What is the meeting point of these approaches, the common ground is the set of traits that Mill says we (are to) use to sort kinds by. This will change depending on the reference class with respect to which sorting happens. But it seems like a useful place to start; and in some sense, this indeed is what scientists working to decipher the meaning (if any) of differences that distinguish races from each other are doing (Chapter 3 Section 3.3).

Let us consider some of the examples Hacking gives for the case of race, and how race may be a statistically significant, meaningful or useful class with respect to particular characteristics.

2.3.1.4 Hacking's Updates Applied to 'Race'

Hacking's new distinctions may not solve the question of whether or not race is a "real" or "superficial" kind. And there are definitely questions surrounding Hacking's formulation of so-called "statistical" notions in a rather peremptory and imprecise manner. But these notions are still, I agree with him, useful. Hacking proclaims:

> Imprecise hand-waving concepts are dangerous when they are given fancy names. They can be put to wholly evil ends. But if we do not give them phony names and are well aware of their imperfections, they can be useful when we need them. We do need this concept [referring to statistical meaningfulness]. Many people –as evidenced by debates going on at the time of this writing, in November of 2004 – are scared of the idea that the traditional list of races employed by traditional racialists might be statistically significant classes. With good reason! (105-106).

Hacking remarks in his introduction:

> Most parts of this essay could have been written last year or next year, but the discussion of naturalism, medicine, and race could only have been written in November of 2004, and may well be out of date by the time this piece is printed (102).

So this is where the potentially expired discussion takes place in Hacking's argument. Examples of claims relating a particular 'race' to particular characteristics of interest are compared: 1.examples of work outside medical research: 1.a. Richard Herrnstein and Charles Murray's discussion of the relation of the class 'African Americans' to 'IQ' in *The Bell Curve*, to 1.b. Philippe Rushton's claim that races are distinguished by many properties in *Race, Evolution and Behavior*; but also, importantly for this thesis, 2. examples of such claims in the biomedical discourse: a. the relation of one's 'risk of Tay-Sachs' to the class 'Ashkenazi Jews', b. the relation of the 'Sickle Cell Anemia (SCA)' trait to the class 'African Americans', c. the relation of 'responding to BiDil' to the class 'African Americans', d. the relation of 'human leukocyte antigens' (HLAs) to 'race'.

First, consider work outside medical research: 1.a. Richard Herrnstein and Charles Murray's discussion of the relation of the class 'African Americans' to 'IQ' in *The Bell Curve* came under severe controversy and criticism but Hacking says it should be distinguished from work of Philippe Rushton. The claim that Richard Herrnstein and Charles Murray made was not that race is a real kind. They did not speak about reality but rather says Hacking about statistical significance –though they might have implied a thesis about statistical meaningfulness too.

> The Bell Curve may show that IQ is a statistically significant characteristic of some American subpopulations, but it is neither meaningful from a biological point of view nor useful for any well-defined purpose (109).

Hacking says this because Herrnstein and Murray did not establish any biological basis for the statistically significant distribution of the trait they recorded. But note his nod to the lack of statistical meaningfulness of the characteristic, 'IQ', "from a biological point of view". This leaves open the possibility that the trait might be meaningful from another, non-biological point of view, such as say an educational one. The quality of educational resources afforded to members of the particular American subpopulations might be another statistically significant characteristic of the group, and provide a causal explanation for the statistical significance of trait

'IQ', in that context. That is, given 'IQ' is indeed a statistically significant characteristic in this situation[82].

Contrast this to Philippe Rushton's claim that races are distinguished by many properties that are important for and indicative of different strengths and weaknesses. Following Mill's definition of real kinds as ones possible to sort by endlessly more features than what we use. These claims of Rushton, regarding the plurality of all these important and not well appreciated distinguishing properties is a claim about the reality of the kind 'race', says Hacking.

> One deplores both Rushton and The Bell Curve, but there is an absolutely fundamental logical difference between what the two assert. Rushton claimed that the races are real Kinds. One imagines that Herrnstein and Murray thought so too, but what they claimed was that the races are statistically significant classes. And they implied that this is statistically meaningful (106).

More interestingly, Hacking considers examples of such claims relating a particular 'race' to particular characteristics of interest in the biomedical discourse (107-109). Take first the relation of one's 'risk of Tay-Sachs' to the class 'Ashkenazi Jews'. In this case we are considering an autosomal recessive hereditary disease, one like sickle-cell anemia, but much less common and much more deadly. In its most common form it affects young children and it is very common in particular population groups one of which is that of people of eastern European Jewish descent. The mutation responsible for the disease causes the lack of an enzyme and causes the accumulation of harmful residues in the brain and nerve tissue leading to a child's mental retardation and ultimately to death. Hacking claims: "Ashkenazi is a valuable geographical, historical, and social classification" (107).

Note that he does not say "biological". He continues:

> It is geographical because it indicates where members of this class, or their near ancestors, came from, namely, eastern Europe. It makes a contrast with Sephardic Jews, whose roots are in Spain. In modern Europe and North America, social differences between the Ashkenazi and Sephardic hardly matter to most people, but they remain significant in North Africa and West Asia. Until further interbreeding makes it totally obsolete, Ashkenazi is a statistically significant and a statistically meaningful class with respect to Tay-Sachs disease (107).

[82] Hacking claims he refutes Herrnstein and Murray's claims in an older essay (footnote 4, 106).

Hacking is careful to distinguish between the social differences between the Ashkenazi and Sephardic Jews, and what characteristics the groups have that may depend on factors like 'where their ancestors came from' or 'interbreeding'. Note that the claim Hacking makes is not about the reality of these categories; only about their statistical significance and meaningfulness, in this context.

So recall:

(HSSC) P1 is a *statistically significant* class with respect to characteristic c if and only if the distribution of c in P1 is significantly different from its distribution in some implied comparison population P2.

(HSMC) P1 is a *statistically meaningful* class with respect to characteristic c if and only if c is statistically significant and its variation between P1 and some implied comparison population P2 is understood in terms of causes.

Following (HSSC) 'Ashkenazi Jews' is a *statistically significant* class with respect to characteristic 'Tay Sachs' as the distribution of 'Tay Sachs' in 'Ashkenazi Jews' is significantly different from its distribution in the implied comparison population 'Sephardic Jews'. Following (HSMC) 'Ashkenazi Jews' is a *statistically meaningful* class with respect to characteristic 'Tay-Sachs' because 'Tay Sachs' is statistically significant and its variation between 'Ashkenazi Jews' and implied comparison population 'Sephardic Jews' is understood in terms of causes.

What about the relation of the 'Sickle Cell Anemia (SCA)' trait to the class 'black', in the context of the U.S.? Hacking notes:

> West African ancestry is an indicator for being a carrier of the sickle-cell anemia trait, which confers some immunity against malaria. This trait was often stigmatized as simply 'black'. In fact, it is primarily West African, although it shows up in Mediterranean populations where malaria was a major selector for survival. The indicator was abused for racial reasons in widespread screening (107).

Hacking does not include this case in the ones he reformulates in terms of statistical concepts but given the importance of the case for Root's argument one might attempt to do so here. The trait we are interested in is I presume 'being a carrier of the sickle-cell anemia trait'. So, the trait is, we read, "primarily West African". Also "West African ancestry is an indicator

for being a carrier of the sickle-cell anemia trait" –where by an 'indicator' I am taking Hacking to mean what was defined as a 'proxy' ((EBP)in Section 2.1.4.1.3). Is this enough to go on? Not so. We are lacking a clearly defined comparison population in this account. That is unless we take Hacking's claim that the trait is "primarily West African" to indicate that the relevant comparison population would be 'non-West Africans'. In that case, we might say that 'West Africans' is a *statistically significant* class with respect to characteristic 'the sickle-cell anemia trait' as the distribution of 'the sickle-cell anemia trait' in 'West Africans' is significantly different from its distribution in the implied comparison population 'non-West Africans'. And similarly for articulating the characteristic's statistical meaningfulness for this population, given the particular comparison class. Note that the claim we established here is not the one Root argues against. The class 'West Africans' may be a subset of the class 'black', so 'West Africans' will be 'black', but 'non-West Africans' need not be 'non-black'.

However, assuming that it is the case that the classes made up of those 'West African and in the U.S.' and those 'non-West African and in the U.S.' are both populations we find in the United States, and further, if we could assume that the class 'West African and Black and in the U.S.' approximates the class 'Black' and the class 'non-West African and Black and in the U.S.' is of a negligible size, then we might proclaim that the class 'black' would too, in this situation, be statistically significant and meaningful –and useful.

Coming to the relation between 'responding to BiDil' and the class 'African Americans', a case that was important for Daar and Singer's argument –which though did indeed expire– Hacking notes:

> Some medications may be less effective and BiDil may be more effective, for African Americans with certain types of heart failure. If so, this is statistically significant and statistically useful for helping patients, but (in my opinion) it is at present not statistically meaningful (109).

Hacking claims that 'race' was a statistically significant class and a statistically useful one in the case of 'responding to BiDil', but not a statistically meaningful one. This is because, as already discussed, there is no clear causal account for why 'responding to BiDil' may be causally linked to the class of 'self-identified African Americans' besides an association between 'a nitric

oxide deficiency' and 'self-identified African Americans'. The case of BiDil explicitly shows that meaningfulness and usefulness can become disjoint.

Recall the two definitions proposed for a *statistically useful* class:

(HSUC) P1 is a *statistically useful* class with respect to characteristic c if and only if *c* can be used as a proxy for a feature of some fairly immediate concern

(HSUC)' P1 is a *statistically useful* class with respect to characteristic c if and only if *P1* can be used as a proxy for c, where c is a statistically significant characteristic of P1 relative to some implied comparison population P2 and of some fairly immediate concern.

Which definition would apply in this case?

If P1='self-identified African Americans' then, what is the comparison class we are considering? It is not specified here. Knowing that Cohn and Carson applied for a new drug approval to the FDA having retrospectively analyzed subpopulation data obtained under other racial classes in their original trials on this substance (Section 2.2.1.5.1), we might say that that in practice there was a comparison population for establishing the claim, and this is 'the set of other U.S. 'races' under which data was collected', say P2='self-identified Whites or Hispanics or Asians or Pacific Islanders', which is different to the class 'not self-identified African Americans'.

Ok, so having specified P1 and P2, it is starting to become apparent that it is the second definition (HSUC)' that applies to this case. P1='self-identified African Americans' was indeed used as a proxy for c= 'responding to BiDil' –or more accurately it was a proxy for c='responding to the hydralazine isosorbide dinatrate combination H/I', at that stage in the testing. Was the characteristic 'responding to BiDil' a statistically significant characteristic, for P1='self-identified African Americans' and P2='self-identified Whites or Hispanics or Asians or Pacific Islanders'? Yes. That is what analyzing subgroup data indicated. Recall:

Given classes P1 and P2 and c, a shared characteristic of these classes,

(HSSc) c is a *statistically significant* characteristic if and only if the distribution of c in P1 is significantly different from its distribution in P2.

The distribution of c= 'responding to BiDil' in P1='self-identified African Americans' was significantly different from its distribution in P2='self-identified Whites or Hispanics or Asians or Pacific Islanders', after retrospective data analysis. This was the claim put to the FDA in demand of a New Drug Approval (NDA).

What about (HSUC)?

(HSUC) P1 is a *statistically useful* class with respect to characteristic c if and only if *c* can be used as a proxy for a feature material to some fairly immediate concern.

This definition is much more open-ended. Under this definition 'self-identified African Americans' would only be a statistically useful class if it could be used as a proxy for a feature material to some immediate concern. This is not that implausibly the case. For example, it could be argued that 'responding to BiDil' is well correlated with one's 'heart-failure outcomes' –this is, if the drug actually works as A-HeFT suggests. This second is a feature material to a fairly immediate concern: a patient's health. So we might thus say that 'self-identified African Americans' is a statistically useful class with respect to 'responding to BiDil'. But you see, it could also be that 'responding to BiDil' is well correlated with 'riding the bus'. This second is a feature material to an advertiser's immediate concern: where to post a BiDil advertisement. So the class 'self-identified African Americans' is statistically useful with respect to 'responding to BiDil', both for those concerned with health outcomes and for those concerned with selling drugs. And this is all without the class 'self-identified African Americans' and the characteristic 'responding to BiDil' being in any way causally related. Though there might be a causal association between 'responding to BiDil' and one's 'heart-failure outcomes' the second association, the one between 'responding to BiDil' and 'riding the bus', will not be meaningful, except in a very tortuous sense of meaningful. Meaningfulness and usefulness are distinct notions.

So, a class can be statistically useful with respect to a characteristic if the characteristic approximates our concerns. The question is what happens when these concerns conflict. What happens when we are concerned for a patient's health, but also, concerned with making a profit. How do we make use of the useful class 'self-identified African Americans'?

Unfortunately, Hacking gives no general direction. He is not necessarily to blame for not giving directions. But he is to blame for not saying why he will not or cannot give general directions.

2.3.2 Should 'race' categories be used to approximate medically interesting human genetic variation?

The answer Hacking gives is: perhaps.

He approves using 'race' in cases where race is a statistically significant, meaningful, and useful class (HLAs matching); and he seems to approve cases where race is a statistically significant and useful class (BiDil). He rejected a case where race is claimed to be merely a statistically significant class (Herrnstein and Murray's claim on IQ), though this rejection could have depended on the fact that this case's claim was mistaken.

Hacking's theoretical frame is in some ways lacking. First, if we are to distinguish classes statistically (as significant, meaningful, or relevant) a distinction between real and superficial kinds can become one of degree. Second, if we can sensibly talk about classes in terms of characteristics we need to have prior knowledge of which characteristics and why these and not others would be relevant to examine; a knowledge which Hacking in part aims to replace with statistical observation and measurement. Third, one must decide what should pass for significant, meaningful and useful. Hacking accepts that –at least– in the case of 'meaningfulness' and 'usefulness' there is no agreed cross-theoretical definition and any claim to the contrary should be contested; but this doesn't mean there are no criteria for judging use and meaning within particular contexts and one may well give some hint as to how to go about establishing these.

Perhaps it is already a dimension of these classes' usefulness, but the ethical worries Hacking has about the illegitimate use of 'race' to oppress people are not explicitly related to the question of using race in biomedicine. Hacking mentions cases of bad research; distinguishes them from 'good' or 'better' research along the statistical dimensions specified. But what about the bare use of these categories; significant, and meaningful, and biomedically useful, or not?

Race should matter for biomedicine, so it should be used in medicine; but race also legitimates oppression.

So what do we do?

Hacking gives no clear instruction.

Hacking (2006) borrows Rabinow's notion of "biosociality" to indicate the distinctiveness of social groupings formed on the basis of (in some cases newly defined, genetic) biological identities. The phenomenon Rabinow describes as biosociality that is that people bond and develop self-conceptions on the basis of biology; his "prophecy" that they will do so on the basis of new genetics classifications (84) means, in Hacking's terms, that we are entering a period of "'making up people' with a vengeance" (94)[83].

This is definitely an interesting prediction to make. But this notion and the neologism that seems to suggest that it is a new phenomenon we are observing here are surrogates for processes that are more general and transcend a distinction such as that rather old and multiply rehearsed one between society and biology.

My remarks have to remain rather cryptic, until an exposition of what I am thinking of in terms of "finding" and "founding" in scientific contexts is offered. But to put it simply –if not trivially– the (social) context that a (biological) thing is in may shape its 'identity', and this is the case for many things other than human biological classes and many contexts other than a social and societal one.

[83] Hacking's views on these topics are discussed in *Historical Ontology* (Hacking 2002) (esp. chapter 2 and 6 on "Making up People"); and parts of *The Social Construction of What?* (Hacking 1999). Hacking also stresses the importance of "interactions" between the classified and a classification in the case of human classifications: "The classifications of [humans in] the social sciences are interactive", by which he means that people respond to the classification they are classed under; classifications in the natural sciences are not subject to the "same type" of interaction (Hacking 1999, 32).

It may be useful to describe this phenomenon at this fineness of grain, and so heuristically of use to give this name to it; but we should not make too much of the singularity of what "new" phenomenon we are observing here lest we lose sight of other dimensions and other ways in which such interactions are and will be already taking place no matter how our biological understanding is couched in social contexts and how they each keep changing.

To use Hacking's terms to raise this worry, biosociality is possibly an interesting concept but when is this indicative of a trait significant, meaningful or useful for medical practitioners to grasp? For example, how is talk of biosocial groups helpful once it comes to figuring out the causal contributions of 'race' to health outcomes?

I think that it is not biosociality *per se* but rather its parsing into the 'bio' and the 'socio' and their possible interactions that is of relevance to biomedical contexts –at least given how disease causation is currently conceived. And I think this is a matter of the metaphysics and notions already in operation in these scientific contexts.

It is how and why a parsing of fuzzy ordinary notions could be happening when these enter scientific contexts of use that *found science* tries to illuminate.

Conclusions to Chapter 2

I considered the work of philosophers of science Michael Root and Ian Hacking and bioethicists Abdallah Daar and Peter Singer. The first purpose of this analysis was to demonstrate that authors use the word "race" to refer to different concepts of race; and to show that in many cases the authors' manifest concept of race is different to their operative one. I argued that these concepts can be organized fairly well if we use the distinction between biorace and sociorace notions of race.

Specifically, I argued that Michael Root conceives of race as sociorace and most likely as socialrace:

1. He claims that race is a construct not a biological kind.
2. He claims that race is a real social kind.

3. He proposes race should be used as a variable in biomedical research but mainly to monitor the effects of racism on health outcomes.
4. He claims that race cannot approximate any medically interesting genetic variation between human groups.

On the other hand, Peter Singer and Abdallah Daar conceive of race as biorace.

1. They take race to be based on continental ancestry.
2. They suggest race is a biological kind.
3. They propose race could be used to track the effects of race-specific genetic variation on health outcomes and to develop race-specific pharmacogenetics.
4. They propose race should be used to track the effects of race-specific genetic variation on health outcomes and to develop pharmacogenetics.

And Ian Hacking has a way to talk about the aptness of race in biomedicine without seeming to rely on a concept of race.

1. He claims race is a superficial kind without ascribing to it a particular nature.
2. He claims race can be statistically significant, useful or meaningful with respect to particular characteristics, in different, biomedical and non-biomedical, contexts of enquiry.
3. What he calls an updated naturalism can justify the use of race in medical research on a case-by-case basis. He suggests race should be used as a variable in biomedical research in cases it is at least statistically significant and useful

Further, philosophers' answers to the normative question I chose to orient this debate were repeatedly found to be lacking:

Should "race" categories be used to approximate medically interesting human genetic variation?

To summarize my response to Root's recommendations –or how missing a biorace race notion leads to trouble

1. race seems to be a good proxy for genetic variation so why not use it to measure potentially significant genetic variations
2. using the term ‘race’ to measure the outcomes of racism varying race-assignment to approximate racism-proneness is confusing and contentious [call it ‘socialrace’; define “racism”, why should we tie up ‘race’ with racism in our definitions, etc.],
3. his claims regarding the biological status of races are not very well-substantiated with empirical evidence though they are empirical claims.

My response to Daar and Singer’s recommendations –or how missing a sociorace race notion leads to trouble can be summed up as follows:

1. race is a category that is local and context specific and it is not used uniformly by the populations of interest
2. using the term ‘race’ to measure the effects of population-specific genetic variation is confusing [name the ‘bioraces’ precisely if you can; define “genetic variation”, why should we tie up ‘race’ with genetic difference in our definitions.]
3. their claims regarding the applicability and usefulness of racial categories for pharmacogenomics research are not well-substantiated with empirical evidence though they are empirical claims. Maintaining that race can be used to identify the individuals of a global population assumes that the same [historically questionable] set of categories can be projected and taken up across the globe and that so doing will be profitable for the pharmaceutical industry.

Finally my response to Hacking –or how missing the distinction between different race notions can hamper Hacking’s argument is:

a. The notions of statistical significance, usefulness and meaningfulness are used to operationalize Mill’s distinction between real and superficial kinds but no clear thresholds are drawn.
b. Hacking talks of the statistical significance, usefulness and meaningfulness of a population having defined the notions in terms of particular characteristics but this can be

> problematic. A choice set of relevant characteristics and populations needs to be defined. This involves prior interests and knowledge about these populations and characteristics. And at the end of the day one must still decide what should pass for significant, useful, and meaningful. Hacking cannot give us a rule of thumb outside a context and he can start by being more precise about what types of characters we should care about in different settings and why.

In a time where there is political interest in addressing the disparate health needs of the various US socioraces the risk of essentializing sociorace categories and taking them for bioraces is great. To avoid conflating the two dimensions along which one's 'race' is thought to shape 'health' the concepts of sociorace and biorace need to be formally distinguished and their causal import needs to be investigated.

Mike Bamshad of the department of Pediatrics and Human Genetics in the University of Utah has suggested that the OMB classifications should be dropped for the government sectors working on health and some other notions be put in their place:

> Referring to 'geographical ancestry' instead of race is an emerging alternative that is both more accurate and less contentious. One way to operationalize this approach is for the National Institutes of Health to change its current requirement to use Office of Management and Budget categories and instead mandate stratification of individuals by self-assessed descriptors of ancestry such as the geographic origin of an individual's parents (e.g., Central Africa, Southeast Asia, Central America), followed by their ethnic identity, and finally the community in which a person resides (Bamshad 2005, 945).

This suggestion seems to be breaking up 'race' into a biological component (ancestry) and a social component (ethnic identity, community). This accords with the thesis defended here: that there are two dimensions along which the notion of 'race' is operative in medicine: a biological and a social one. Of course Bamshad is not here interested in replacing race. He is interested in specifying what about race might be relevant to biomedical outcomes. So, the notions of "geographical ancestry" and the two of "ethnic identity" and "residential community" can I think be respectively understood as ways to articulate, and so then to operationalize or approximate different biorace and sociorace notions of race that may be biomedically important to measure.

Besides the aptness of these notions as opposed to other ones, as good proxies for the purposes intended there are questions of when –at what level of institutional structure and in what biomedical context such a distinction would be relevant and useful. First of all: what do we do with data already obtained along race/ethnicity categories? Should we translate these into the new categories? When should we do so and how? Might there be reason to keep using the term “race”, vague as it is, to hold sight of social pressures specific to it, or would information along the categories of ethnic identity and community always be collected along with that of geographical ancestry, even in a biomedical genetics context?

These are questions that I cannot answer but questions that need to be asked and answered by the experts in the particular situation in question. Making the process of ‘proxy-picking’ transparent allows the evaluation of the proxies and for that you need a notion of what exactly is being approximated: what I argued are sociorace and biorace notions of ‘race’.

It is only by explicitly acknowledging the complexity of notions like ‘race’ that we stand a chance of understanding and combating the social and biological ills that make people unhealthy.

Chapter 3
DIAGNOSIS

Found Science: Founding 'Race' in Biomedical Science

This chapter makes the positive proposal of this dissertation: it offers an understanding of why this multi-use of "race" does and should happen in the process of studying ordinary race in biomedical research.

I argue that a process is being followed by practitioners in these fields which enables studying ordinary race concepts in science but comes at the cost of some transfiguration to these concepts; enough 'transfiguration' to render them distinct from their ordinary forms but also necessary to extricate them from their ordinary intentional contexts and allow them once embedded in science to function in ways extraordinary and thereupon unforeseen.

With attention to this process and appreciation of its intricacy and value further judgment can be made about the costs and benefits of this process for the broader, ordinary, concerns that concern a useful science.

3.1 Defining Found Science: A Process for Founding Common Concepts in Science[84]

Found science describes how a common concept becomes scientific by analogy to how an ordinary object becomes art. Found science claims that

> common, "non-scientific" concepts can become part of science practice, if they can be transfigured appropriately.

[84] The chapter is based on a talk prepared for Nottingham University's Institute for Science and Society and was subsequently presented at LSE and at UCL. Thanks are owed to Hakan Schechinelgin, Conrad Heilmann and Damien Fennell for their questions, to Donald Gillies for his further suggestions and to Nancy Cartwright for support which took many forms.

This thesis is more general than what is needed to describe the case of race concepts' use in biomedical science: its target is specified as "science" as opposed to particular genetics or epidemiological contexts[85] and its argument could be an ordinary concept other than race.

I offer it here in this form knowing that this case study of 'race' concepts' use is doomed to fall short of confirming it. I do so not as a fugue for the imagination but as an indication of the scope of interesting questions that may be asked within the proposed project, were it to lead to more research.

The purpose of this idiosyncratic story is to describe something seemingly banal: the use of what looks to be the same concept in more than one way and what consequences such systematic usage(s) may have. In the case of race and its use in biomedicine, we would be talking about the same term, "race" as referring to the seeming 'one' concept, and its use in different biomedical contexts as honing in the ordinary concept of race to 'sociorace' or 'biorace' notions of race. But for now, let us consider what finding and founding ordinary concepts in science would look like.

3.1.1 The Thesis of Found Science

Found science makes one central claim:

> An available, "non-scientific" concept may become a part of science, if –but only if –it is transfigured appropriately.

The word "transfiguration" is chosen on purpose. It describes a change in shape, a transformation; but it has a negative connotation. A transfiguration changes a thing's 'normal' form to a shape that is unusual and often unappealing. This transformation can be damaging but it may unleash powers the subject would not otherwise have[86].

What is the question that "found science" answers?

---

[85] Of course the process can be articulated in more general ways, still; as talking about installations of entities in contexts of practices other than scientific ones. The apparatus that supports the more general articulation needs to be significantly further theoretically articulated and evaluated than fitting what numerous simple examples I can fit to it (cf. Appendix 4).

[86] Plus Arthur Danto titles his book on found art *The Transfiguration of the Commonplace*, Danto (1981).

Heidegger claimed that the metaphysics of an age seeps through its cultural productions and is reflected in their "essential shape" (Heidegger 1950 [2002], 57) so perhaps an analogy between what *found science* does and a recent film form is not too vulgar: I am referring to the "prequel".

The first "prequels" I heard called such were *Episodes i, ii, iii* of the science fiction film trilogy *Star Wars. Episodes iv, v, vi* were filmed in the 1970s and recounted how a lone Jedi knight, Luke Skywalker, led a diminishing order of rebels to overthrow an oppressive empire ministered by Lord Darth Vader, or Anakin Skywalker, Luke's thought to be dead –but turned evil– father. The story ends happily in the 1970s films with the rebels overthrowing the empire and Luke making peace with his father's ghost. So you'd think that was that: What is there to say after a happy ever after?

Something, I suppose. *Episodes i, ii, iii* were "prequels" filmed in the 2000s and made to fill in the story of how Anakin Skywalker turned into Darth Vader.

It is this more "boring" type of work that found science does; it fills in a gap and a gap of origins it is.

How does an ordinary concept become a part of "science" and what is this "science" that is so formed?

Found science is a "prequel" for what well-established theories in philosophy of science discuss as the contents, forms and values shaping different but well-defined "sciences".

Before science becomes "normal" or "anomalous", before it gets to be "programmatic", before it can progress or regress, before it gets to be value-free or value-laden, right or wrong; before all of that, science must be found –and it stands to be founded.

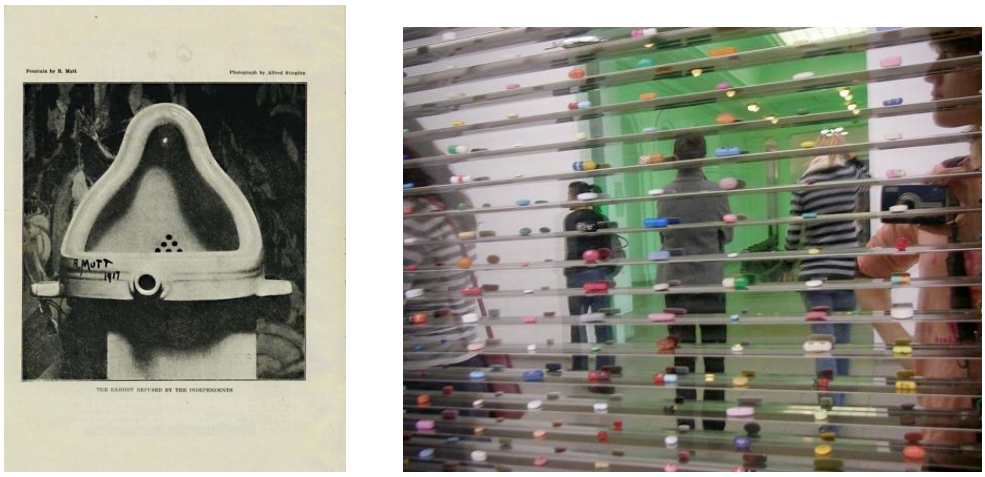
I claim that found science is 1. *science* and that it is 2. *topical* science. This is because *found* science only becomes found *science* when it's *founded* properly in a scientific context; and it cannot get founded unless it is in the first place *found*.

3.1.2 Defining Found Science By Analogy to Found Art

I define "found science" by analogy to "found art". So let me ensure that the familiar arm of this analogy is indeed familiar.

Found art has been around for approximately 100 years. The first pieces of what is now called found art were Marcel Duchamp's readymades (or *objets trouvés*)[87]. Readymades influence art through the 20th century. They become especially visible with the Young British Artists in the early 1990s who count Damien Hirst, Tracy Emin, and others in their ranks. Besides the visual arts, found art influenced the performance arts and music, notably the compositions of John Cage[88].

Below are two typical examples of found art. On the left is one of the first instances of found art, *Fountain*, by Marcel Duchamp, first exhibited in 1917. On the right is a more recent piece of found art, *The Void*, by Damien Hirst, first exhibited in 2000 and photographed below in Berlin's Hamburger Bahnhof (Figure 3).



Fountain, Marcel Duchamp (1917) *The Void*, Damien Hirst (2000)

Figure 3: Examples of Found Art

What I'm interested in is what is radical about found art. This is the claim –put to practice– that a common object, an object seemingly foreign to art and even *counter* to our

[87] Cf. the biography of Cros (2006) for more on the fascinating character of Duchamp.

[88] Cage uses found sound and found rules. See Cage's use of chance to compose a piece but also as an element in the piece's performance. For example Cage used the *I Ching* to compose *Music of Changes* , solo piano (1951). His *Imaginary Landscape No. 4* (1951) is a piece for twelve radio receivers tuned to specific frequencies, volumes, etc. but with no control on what is being broadcast during the performance.

expectations of art, like a dirty urinal, or a disturbing pill, or the sound of biting an apple, can *become* art[89]. How? By being placed in a context which *makes it* art.

You may notice that the urinal in *Fountain* is turned upside down and inscribed ('R. Mutt') and know that the pills displayed in *The Void* are replicas of pills. Found objects are not displayed exactly as found, nor are these objects indeed always "found". Still, the idea is that any seemingly banal object has a claim to being art:

"Anything can be art" –for example a shark, placed in a tank of formaldehyde (Figure 3),

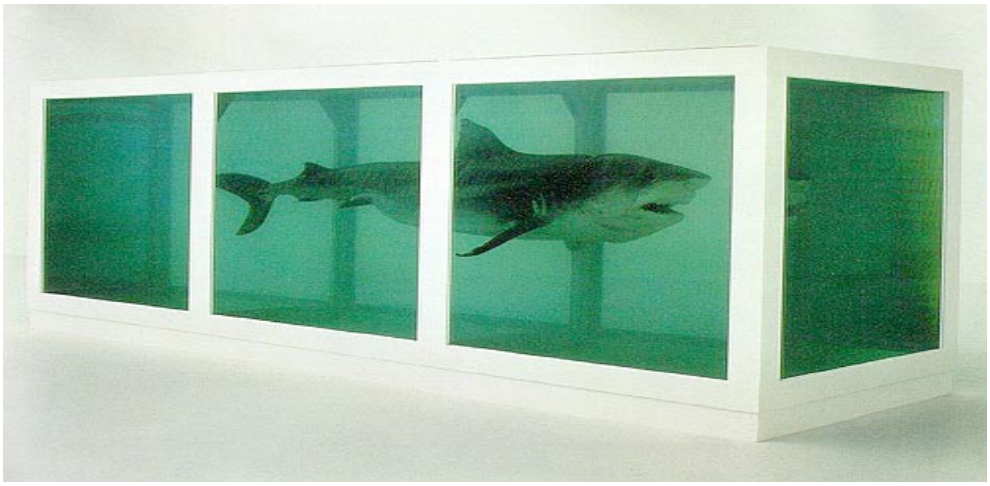


Figure 4: A Founded Shark

nicely titled (*The Impossibility of Death in the Mind of Someone Living*) and appropriately authored (*by Damien Hirst* in 1991)– "if it is placed in the artistic context"[90].

But what is it to "*place* something *in* an artistic *context*"?

To my understanding, two motions are involved in placing something somewhere: 1. a picking up of the thing, out of a context, and 2. a placing of the thing, in a context.

What kind of a context is an *artistic* context?

It is a context wherein the thing can *sit, as art*!

3.1.2.1 Founding a Shark in Art

Let us look at the piece pictured above (Figure 4). How is a shark found art?

[89] Cf. Danto (1997) 84 quoting Marcel Duchamp: "When I discovered readymades, I thought to discourage aesthetics. … I threw the bottle rack and the urinal in their faces as a challenge, and now they admire them for their aesthetic beauty".
[90] Cf. http://arted.osu.edu/160/02_Intro.php [Last accessed: 09/07/07]

First, the shark, a ready-made thing, is *found*. It might just be found *by* an artist –as was probably the case with this shark. But it could be that the thing is picked up by an artistic tool or a process –for example, someone mopping the floor, is recorded by Papalexandri-Alexandri's sound-recording in *Sta Skalia (On The Steps)* (2003)[91].

Second, the shark is *founded*. It is founded in two sorts of context: First, the shark is founded in a *physical, spatial context* that carries the art stamp. It is placed in green formaldehyde, inside a white tank, in an art studio, or an art gallery and so on. Second, the object is founded in a *context of interests* that carry the art stamp: the interest of the artist in the object, the curator's interest in the object, the interest of the audience members in it, etc.

These two types of context interact. Artistic interests shape the physical context of the object – and the physical context of the object affects what interests it attracts. Hirst wants the tank to be white so he paints it white, the curator likes the piece and so he installs it in a prominent space in a gallery, etc. And reversely, artistic interests are shaped by the physical context of the object: The curator likes the white tank, the buyer develops an interest after seeing the object displayed in a famous gallery, etc[92]. Note that a lack of interest in some property of the object can also carry the stamp of art; for example, a lack of interest in the life of the shark distinguishes it from an aquarium[93].

Both of these contexts, the physical, spatial context and the context of interests, found the object *as art*. And once the object is *founded* as art the whole piece *functions as art*: it can be found again, *as art*.

And so it is that an ordinary fish comes to be art. It loses some of its capacities (its capacity to swim for one) while others (its capacity to scare human audiences perhaps) are exercised, and in fact "trained", exercised again and again, in different art museums, systematically and preserved in the documents of the art world for generations to come.

---

[91] I am referring to the piece *Sta Skalia (On The Steps)* for tape, by Marianthi Papalexandri-Alexandri (2003)

[92] This expands on what Schaeffer (2000) 289 calls the inseparability of a "processual intentionality" and "the artist's encounter with the medium worked in".

[93] These interests might follow a canon as McAllister (1996) notes, but the content of the canon will differ in time and space. Doppelt's (ms) account of epistemic values may be relevant in these embeddings too.

This shark is no longer a fish –nor *fishy*. But it is dead.

And note: This is not a story about the artistic *merit* of the shark-installation. It is a story about its *artistic* merit. The piece could be good or bad art by what standards we care to consider but it will be good or bad *ART*.

This is also not the real story, nor a complete story. But it is the right story for what I want to do, which is abstract from it.

3.1.2.2 Founding 'Science'?

How can this be at all relevant to *science*?

First, take this association between words: I have been using two forms of 'found', the past participle of the verb 'to find' and the verb 'to found'. 'Found art' refers to art that *is found* but also to the process needed *to found* an object as art. So the found art object is found in two ways: 1. It is discovered –it is found. Rather than created the object appears to be readymade or already available, 2. The object is supported –it is founded. The object is installed in contexts of physical, discursive, social or other type of space and it is held there by corresponding contexts of interests –interests in values of different sorts, but with the object at their focus.

That's where the resonance with the notions of discovery and justification stops. I used the terms 'discovery' and 'justification' here because they are familiar; not because they are exactly what I mean[94]. 'Discovery' is not an apt concept for what I am describing here: it is not a "dis-sing" of a covering, an "unveiling" or such that I am interested in. No –There is no such proclamation made here. What is found is not scientific unless it is founded. And founding does not consist in mere logical justification –nor brute historical imposition.

'Finding' and 'founding' are broader notions than discovery and justification and are meant to describe all sorts of actions that result in concepts' placement in the contexts of science; not only steps in scientific theory formation. Finding and founding form two actions of a process

[94] Various types of a context distinction between discovery and justification along with their shortcomings are reviewed in Hoyningen-Huene (1987). The distinction can be one between: a. processes, b. the process of discovery and the method of justification, c. the empirical process of discovery and the logical testing of justification, d. empirical and logical academic disciplines, e. types of questions.

that must be run repeatedly before a concept is properly founded in a context and run recursively in different contexts. It is an interaction of space and interests where the two steps happen that sustains a concept as relevant to science.

In that sense discovery and justification morph into each other and what is "discovered" *as* science, can –if you follow this frame- only justifiably be deemed recovered and re-covered, in science[95].

These two types of action, finding and founding (using tools), I suggest can turn an ordinary concept into a "scientific" one.

3.1.2.3 Founding a Concept in Science

Now, given a concept and a context of science, when would a concept be 'found science'?

First it would be a concept that is available but taken to be "non-scientific", "ordinary" or "other" in that context of science. This concept would be already available i.e. already used or functioning in a context other than this particular one of science but accessible to agents/processes working in the scientific context.

Next, embedding a concept in a context of science would involve two types of action:

1. First, we would "find", "pick up" and "pick out" the concept from its original context. It would be the case with most ordinary concepts that a user would already be in possession of the concept. So finding a concept need not be as unlikely as finding a fish.
2. Second we would found, "install" a concept in target scientific contexts. Founding actions would be actions like articulating in scientific terms, using in scientific arguments, fitting to scientific theory, discussing with colleagues, operationalizing with a view to measurement, etc. For example a scientist might define the term associated with an ordinary concept in the terms specific to her disciplinary context or aims; or she might

[95] This understanding is in line with the etymology of the Greek word for 'discovery', "anakalypsis": This is a compound of 'ana' + 'kalypto' or 're' + 'cover'.

physically manipulate the object a common concept is supposed to specify along the precepts of her discipline.

Third, this two-part process for embedding the concept would happen in contexts of two types:

a. In *spatial* contexts stamped by science: By these I mean both physical actual spaces but also spaces of ideas, like the particular ontologies and theories of scientific contexts. These could be physical, discursive, social spaces of science like, science theories, science labs, or science journals or science conferences. They would be answers to the question: Where is the concept used? In the lab, in *Science*, at the NIH, etc…

b. In contexts of *interests* stamped by science. These would include *interests of* science producers and users *as* science producers and users. They would be expressible by questions. They could also be guided by "values" one might attribute to this particular context.

Further, contexts of interests in a concept would determine the "spatial" context it ends up in and vice versa what space the concept is in would shape the interests it attracts. For example, a biologist interested in the concept 'competition' may decide to publish his research in a biology journal. Biologists reading the biology journal might develop a biological interest in 'competition' because they read about it there.

The concept would be founded *as science* in both types of context. Once founded as science, the concept would *work as science*: Which means, it could be found again in the context of science, *as* *science*.

3.1.3 Three Admonitions

Found Science comes with three warnings.

1. First, finding is necessary but not sufficient for founding.

Finding a shark in a context of art is not sufficient for founding it there as art. Hirst could have thought 'shark'! But got a craving for shark steak, or made a joke about a "found shark", or forgotten about it. A found shark is not art, unless it's founded in a context of art.

Same for a science context: Finding a concept in a context of science is necessary but not sufficient for founding it there and being able to subsequently find it there, *as* science. Finding a 'rat' in a lab, is not finding a 'lab rat'.

2. Second admonition: As mentioned already the story I told was incomplete. But there is a particular part in the story that needs to be filled in. And it's this: We use tools. The shark is transported from the fishery in a car; it is loaded in a tank with a crane; it is preserved with formaldehyde, etc. And tools are also used to place the shark context of interests of art: Hirst becomes inspired after seeing a movie about sharks, the shark is made appealing to the buyer by hearing of Hirst's sales, an art aficionado gets interested in the exhibit after reading the paper, etc. Some tools are particular to art. Many are not.

Further, tools work by "picking up" the shark in a way that preserves features we want to "place" in the new context. For example, a van, a crane and formaldehyde work for Hirst because they preserve the form of the shark. Similarly for a scientific context: tools will be used to select and preserve what features of a concept we care to and are able to study in the scientific context.

Although I think this point about tools is important for the ultimate development of the idea of found science, I don't have enough worked out about it at this point to elaborate it here. I will call "finding" tools tools that are used to "pick up" a concept out of either a spatial context or a context of interests. Of these finding tools some will be apt for "placing" it in either a spatial context or a context of interests of science. I will call these "founding" tools. Founding tools are finding tools whose function is founding a concept in science.

3. Finally, found science is <u>not</u> a story about the scientific <u>merit</u> of an entity found as science. It is a story about its <u>scientific</u> merit.

This process for embedding an ordinary concept in science doesn't come with the mandate that what work the concept does in there is of any value. Found science can be good or bad science by whatever standards we care to consider.

But it will be good or bad ***SCIENCE***.

3.1.4 What's the Point of this Analogy?

This is admittedly a funny analogy. But it drives home a point central to the thesis of found science:

> "Science" is in principle as *intrusive* to an ordinary concept as formaldehyde is to a shark.

Though many concepts can in principle be part of science, this is conditional on their appropriate transfiguration. Just like a found 'art object' is not properly speaking a 'shark' so a found 'antibiotic' is not properly speaking a 'mould'[96].

This is possibly sad news: founded scientific concepts will differ from ordinary ones if properly founded in a scientific context. And they will differ from each other if founded in different scientific contexts. As a result, extra work will be needed to bridge these contexts and relate our notions of and uses of the founded concepts to each other and to our original ordinary understandings. Seeing found science concepts as distinct from ordinary ones matters for understanding what science is and *can be* doing.

But there is possibly good news too. Installation in a particular scientific context can make manifest different uses that an ordinary concept may have. One could understand this as a freeing up of a power 'in the concept' so to speak; for example, one might say that it was always part of the Concept of a 'shark' that in some cases it would specify a thing usable as an art exhibit: the shark just had to come across the relevant context of use for this possibility to become manifest. (The good news is that a founded concept holds that possibility visible.) Ergo what I called a founded concept is still identifiably a transfiguration of the original concept.

---

[96] I am referring to the case of isolating penicillin from the (rare) mould penicillum notatum. Cf Gillies (1993) 39-48, (2006)

Alternatively, one might think of a concept when transfigured for a new set of practices as a new "found" concept. It depends on one's metaphysics. I have tried to choose a terminology that makes both possibilities visible. But in any case the concept would no longer be 'ordinary', it would be extraordinary and 'scientific'.

3.1.5 What's the Point of "Finding and Founding" Science?

To paraphrase Arthur C. Danto on [art][97]:

"Overcoming the gap between [science] and life".

Found science is science that is *topical*; it is science of the things commonly found and commonly interesting.

It allows us to study in science the things we care about in life. But it comes with the responsibility of 1. founding the things we find while respecting things themselves and 2. allowing them to be found again, in life, *as* science.

But more specifically, as was stated in Tools (Section 1.4), the point of *found science* is to use it.

This story helped me disentangle the divergences between what would be manifest concepts of race and operative ones in discussions of this topic; and further the divergence between these manifest concepts and some one ordinary one. How?

The ordinary concept 'race' is found and is being founded in different contexts differently, in accordance with the contexts' ontologies and interests, and it is being used to effect different goals in each context. These are though still oriented by usage of the same ordinary word "race". When authors interpret results from particular disciplinary domains secured under "race" as referring to the ordinary concept 'race', this leads them into trouble.

It is to this possibility that I turn next. But before going further let me standardize my usage of some of these terms.

---

[97] Danto (2000)

3.1.6 Terminology for “found”, “founded”, “finding” and “founding” Science

1. Grammar

I am using two verbs in passive voice to talk about what happens to concepts when embedded in scientific contexts. When I say that ‘concept x is found’, I am using the verb ‘to find’ in passive voice –i.e. the verb ‘to be’ followed by the past participle of the verb ‘to find’ which is ‘found’.

Similarly when I say that ‘concept x is founded’, I use the verb ‘to found’ in passive voice –i.e. the verb ‘to be’ followed by the past participle of ‘to found’, ‘founded’.

This sometimes gets confusing as I am also using the past participle ‘found’ as an adjective. So in the phrase ‘x is a found y’, ‘found’ is operating as an adjective describing y (and a homonym –the same word actually– for the verb ‘to found’), though it is the past participle of the verb ‘to find’.

The present participles of ‘to find’ and ‘to found’ are ‘finding’ and ‘founding’ respectively and what is used to form the present continuous tense in the sentences ‘x is finding science’ and ‘y is founding science’.

I similarly use one of this present participles as an adjective: in the formulation ‘z is a founding tool for science’, founding is the present participle of the verb to found. Though luckily this doesn’t sound or look like a version of the verb ‘to find’.

2. Definitions

Although regimenting my discourse in the following way may lose some of the nice metaphorical allusions from the found art domain, I propose to talk like this:

1. A common concept is *found* (vis-à-vis found science) when it is introduced into a scientific domain.
2. A common concept is *founded* in that scientific domain when it is transfigured in appropriate ways to support the uses to which it is put there.

Recall that the definitions above specify start and end points of a process. But it is a process that is running repeatedly and at different scales within a context. We know when the process is done because at that point the concept is one that can be found in that scientific context as a scientific one with no need to transfigure it. So,

3. Once a common concept is founded in a scientific domain, it will be possible to find it in that context as a scientific concept.

And to remind you I also use 'found' as an adjective as follows:

4. An x is a found y when x is used as if it were a y. This is equivalent to saying that x is founded in a context of y-use.

With respect to a common concept then:

5. A common concept is a found scientific concept, when the common concept is put to the uses that the scientific concept is put to. This is equivalent to saying that the common concept is founded in a context of use of scientific concepts or a scientific context.

3.2 Founding an Ordinary Concept in Biomedical Research: 'Race'

Describing the use of race concepts in U.S. biomedical genomics as a case of found science is in no way as easy as applying a recipe (which is often quite hard to do in the first place). But these would be the ingredients of the recipe of *found science* in the case of embedding race concepts in genetics. I outline them here but use three studies (Pritchard et al 2000, Rosenberg et al 2002 and Tang et al 2005) to describe some particular steps in such a process.

In this case, ordinary, non-scientific concepts are non-genetic or what will be termed "subjective" concepts of race (section 3.3.1.1.2). These concepts are taken to be "inappropriate" or "non-scientific" within the particular context of science practice but they are "already-available": They are explicitly acknowledged by the OMB as representing "political-social constructs" (and I take it that most life scientists would agree with the above statement) and

already used according to the rules of a context "other than" this particular one of science practice, namely, demography[98]. As examined in two case studies these race concepts are:

1. 'Self-identified ancestry' in Rosenberg et al. 2002,
2. 'Self-identified race/ethnicity' or 'SIRE' in Tang et al 2005.

Next, two types of action are sequentially applied to the non-scientific concept to found it in a scientific context:

1. Finding = "picking up" and "picking out" concepts

----selecting from a context e.g.

-reading samples labels inscribed with the ordinary word

-detecting a distribution of results evoking the concept,

And 2. Founding = "installing" a concept

----placing in a context e.g.

-situating objects thus labeled in lab space

-articulating in terms of a genetic model

-developing measures for it

-relating it with past data

These actions will be examined in more detail in the context of the specific genetics studies.

Third race concepts are installed and sustained in two types of scientific contexts: (a) spatial contexts, (b) contexts of interests. (a) Spatial contexts are the physical, discursive (theoretical) and social spaces where science work on "race" happens. These are spaces "stamped" by science and determined by an answer to: "Where is it?" –In the NIH health directive, on the swab's label, in the lab, in the model, in the algorithm, etc. (b) Contexts of interests in race; who and why they are interested in the category influences what spaces a concept ends up in. These interests are expressible by questions such as: "Is this a population?", "Is this a biologically meaningful pattern?", "Is this a genetic risk-factor?", "Is there medically interesting genetic variation associated with race and ethnicity classes?", etc. These two types of

[98] These are arguably founded race concepts; categories used in the US census and so race concepts founded in the context of demography (cf. Appendix 3 for an argument).

context interact: the fact that a race concept is in epidemiological data influences interests in the concept and vice versa; the fact that a race concept is interesting for geneticists influences its place in (genetic) epidemiology.

Finally, tools are used to perform the steps of finding and founding an ordinary concept in a scientific context. In this case this is computer program STRUCTURE developed by Pritchard et al 2000. It is used to "pick up" a concept out of its ordinary context and used to "install" it in a genetics context.

So, I propose that the following embeddings are taking place:

1. 'Self-identified ancestry' is founded in genetics; it is a found 'genetic ancestry' in Rosenberg et al. 2002,
2. 'Self-identified OMB race/ethnicity' is founded in genetics; it is a found 'genetic ancestry' in Tang et al 2005.

(A study of how race concepts get founded in demography is in Appendix 3 and should be the topic of further historical research. These founded concepts are found census categories; but they are also found to measure people's health outcomes.)

3.2.1 Finding 'Race' in Biomedical Science

"Race" and "ethnicity" variables are currently being used in different biomedical domains in the U.S. Epidemiologists and clinical researchers measure health outcomes stratified by 'race/ethnicity'. Population geneticists and genetic epidemiologists examine the medical significance of genetic variation between 'self-identified race/ethnicity' (or SIRE for short). The Department of Health and Human Services caters to the health needs of different minority populations differently. Drug companies are catering to pharmaceutical needs of particular racial groups. Doctors are prescribing treatment according to race. And patients use race to identify themselves, when seeking treatment.

Below is an excerpt from the OMB's revisions to the 1977 standards[99]. It neatly encapsulates how OMB race and ethnicity standards are officially found in OMB discourse and are expected to be founded in biomedical research:

> The categories that were developed represent a political-social construct designed to be used in the collection of data on the race and ethnicity of major broad population groups in this country, and are not anthropologically or scientifically based.
>
> Next sentence:
>
> The standards are used not only in the decennial census (which provides the "denominator" for many measures), but also in household surveys, on administrative forms (e.g., school registration and mortgage lending applications), and in medical and other research[100].

Categories that are explicitly non-scientific, "not anthropologically or scientifically based", are (to be) used in scientific research. OMB race/ethnicity standards are put in the plates of, and picked up by scientists who use them to stratify their measurements. A ready-made classification, used to count populations in the context of demography, is expected to be embedded in the domains of science[101]. Section 3.3 examines how such an embedding happens in the case of genetics.

3.3 Founding 'Self-Identified Race/Ethnicity' in Genetics

Mandating the use of OMB race and ethnicity as proxies for the genetic structure of the U.S. population would be a policy decision with an identifiable economic and political impact. This analysis deals with a very specific step in a small part of the whole story of how what look to be common categories may legitimately come to be embedded in genetics.

---

[99] See: http://www.whitehouse.gov/OMB/fedreg/directive_15.html

[100] See: http://www.whitehouse.gov/OMB/fedreg/directive_15.html

[101] By 1997 demands for the revision of race/ethnicity standards are explicitly propelled by an interest in race not JUST as a social-political construct but also as a category which measures health disparities. For details see: http://www.whitehouse.gov/OMB/fedreg/directive_15.html Appendix 2, Report to the Office of Management and Budget on the Review of Statistical Policy Directive No. 15, Section 3.5.1.1). My own first look at how ordinary race concepts are founded in demography is cited as Appendix 3 in this dissertation.

Despite the specificity of this analysis non-genetic contextual influences are manifested on even this level of specialized genetics research. It is in light of science as practiced in a social and political context that I write this analysis and it is back to such contexts that I am led to look.

3.3.1 Getting a Founding Tool

It seems that ordinary race/ethnicity population classes can approximate human genetic structure. It is claimed (1) that human genetic structure corresponds to 'self-identified ancestry' by Rosenberg et al. (2002) and (2) that human genetic structure in the U.S. is well approximated by 'self-identified race/ethnicity' categories, in the U.S., by Tang et al. (2005).

Through this dissertation I have made references to these two recent genetics studies. Abdallah Daar and Peter Singer (2005) (among others) made reference to the earlier one as evidence for the existence of medically interesting human genetic variation that corresponds to race. And I used these results myself to challenge Michael Root's claim that race categories are bad proxies for human genetic variation.

It is now time to examine exactly how these studies' claims are secured.

I resist comment on the correctness of these results. This is because I care about their form and the fitness of their form for genetics. I will examine how these two inferences are drawn, in sequence. And I will do so by first examining a tool that they both use to draw these inferences.

I argue that this tool is designed to be used as a tool for founding race concepts into genetics. That is, STRUCTURE is to be used to "pick up" common subjective classes preserving what features of these classes may be relevant to genetics, and it is designed to bring them in a context of use and interest to genetics.

3.3.1.1 Tool STRUCTURE

What I propose is designed to be used as a founding tool for a genetics context is the computer program STRUCTURE.

STRUCTURE was developed at the department of statistics of Oxford University by Jonathan Pritchard, Matthew Stephens and Peter Donnelly. My analysis is based on the paper

detailing their earliest work on the program titled “Inference of Population Structure Using Multilocus Genotype Data” published in *Genetics* in June, 2000. Since this work more versions of the program have been developed. But what is of interest to me here doesn’t change with the ensuing versions. I do not assess the correctness of results obtained using STRUCTURE by some biological or genetics standard. I rather focus on 1. the form of STRUCTURE and 2.the method by which STRUCTURE is used and ask why this tool is appropriate for the use it is put to in this context.

3.3.1.1.1 The Tool’s Mechanism

Like a tool, STRUCTURE works because of its design and like a tool it is designed to work in different contexts and perform more or less the same work.

STRUCTURE runs –and runs according to- a mathematical algorithm. An algorithm is a mathematical recipe which gets its ingredients from the data that the program is given. STRUCTURE’s algorithm is a clustering algorithm: it clusters data into a pre-specified number of clusters according to a predefined mathematical model. More precisely, it clusters genetic data into a pre-specified number of clusters according to a predefined mathematical, genetic model.

In simplified terms the algorithm works as follows[102]. A parametric model of K populations is assumed to apply to the data that the algorithm is fed. Solving a problem involves assuming there is a solution. So, the data obtained from each individual sampled is assumed to originate in one or more of the K populations. And each of the K populations is assumed to be characterized by a set of allele frequencies at each DNA locus. Observations from each cluster are random draws from STRUCTURE’s parametric model and STRUCTURE infers the parameters specifying the model at the same time as it infers individuals’ cluster membership. Inference of cluster membership is made using Bayesian methods.

The goal is to sort genetic data obtained from pre-identified individuals according to some (unknown) genetic rule that transcends their (known) individual origins. Though we don’t

[102] Pritchard et al. (2000) describe this briefly in their abstract and introductory discussion (945,946) but the discussion of ‘Models and Methods’ 947-949 gives a more exact formulation.

know this rule we posit the existence of features that (we expect/ believe, etc.) rule-following will exhibit if in place. If these features are picked up in the samples by our sorting method then rule-following and the rule are confirmed at the same time. Bayesian methods of statistical inference allow us to include prior information we may have about the data's structure in the guise of prior probabilities which we attach to certain distributions of results. By pre-weighing our end results this way we get a chance to control for known imbalances in our samples.

So, hard work originating from within a science discipline, following particular scientific precepts and serving particular scientific aims has gone into developing STRUCTURE. And my question is: Why?

Why go to this trouble?

Why infer population structure *this* way?

3.3.1.1.2 Two Projected Uses

Pritchard et al. (2000) answer.

> The definition of populations is typically *subjective*, based, for example, on linguistic, cultural, or physical characters, as well as the geographic location of sampled individuals (945).

Here the authors make reference to common ways of sorting populations for genetic sampling; these would be methods followed by the pioneers of population genetics such as Cavalli-Sforza who would travel to far off places around the world and test groups of people from different locales (cf. Appendix 1, A1.3).

Pritchard et al propose another method to get at genetic structure:

> This subjective approach is usually a sensible way of incorporating diverse types of information. However, it may be difficult to know whether a given assignment of individuals to populations based on these subjective criteria represents a natural assignment in genetic terms, and it would be useful to be able to *confirm* that subjective classifications are consistent with genetic information and *hence* appropriate for studying the questions of interest (945). [Emphases added]

And they repeat their reasons:

> Even in situations where there is *nongenetic* information that can be used to define populations, it may be useful to use the approach developed here to

> *ensure* that populations defined on an extrinsic basis reflect the underlying genetic structure (955). [Emphases added]

The proposed program, STRUCTURE, is hoped to be useful in two ways. First, STRUCTURE can be used to *confirm* or *ensure* that assignments of individuals to populations based on non-genetic information like geographical origin, physical features or language represent "a natural assignment in genetic terms" or "reflect the underlying genetic structure". Second, STRUCTURE can be useful for *inferring* that these "subjective classifications" are appropriate for studying questions of interest to genetics.

So these are two projected uses for STRUCTURE:

a. STRUCTURE can be used to check if "subjective" assignments *naturally* make sense in genetic terms.
b. STRUCTURE can be used to confirm that "subjective" classes are apt for answering questions of interest to genetics.

Geneticists should infer population structure *this* way to confirm that they are talking in *genetically* sensible terms when using ordinary populations to answer questions of interest to genetics.

Nowhere is it argued that populations defined using STRUCTURE are of absolutely more or less worth, more or less essence, more or less reality than assignments derived using other methods. This way of inferring population structure is nowhere claimed to be "better" *simpliciter*; it is claimed to be "better for" the purposes of genetics.

Of course calling non-genetic definitions "subjective" seems to suggest that a presumption is made with respect the epistemic status of non-genetic markers. And referring to genetic structure as "underlying" these non-genetic features hints at further assumptions about what type of structure causes what other.

These assumptions may be what they may.

The claim I examine here is the claim that this is a method for ensuring the aptness of ordinary categories for the context of interests and use of genetics; whether what those working in these contexts take to be appropriate for their context relies on prior ontological assumptions

and how is a matter for another inquiry. So my question is –forget truth– what about this method gets you appropriateness and how?

3.3.1.2 Why Would Using STRUCTURE Make *Genetic* Sense of Data? I.e. Why is it A Founding Tool for the Context of Genetics?

How *can* STRUCTURE be better able to pick out genetic structure than those "subjective" methods?

STRUCTURE was designed to do so.

The method was designed to

1. *sort* through data, which originates like linguistic features or physical characters in a "subjective", ordinary context, according to rules particular to the context of genetics and
2. to *articulate* the data in terms particular to genetics

where (1) and (2) happen at the same time.

Using this clustering algorithm succeeds in abstracting data from their original contexts and structuring them according to a model already formulated according to the precepts of genetics as one which renders the data articulate in the terms of genetics.

Allow me to explain this more graphically. Things are multi-capable (and we are resourceful). I gave my example already. My pen-cap has an arm which is supposed to help stick the pen in different places like your shirt pocket. I don't wear shirts but I have a fringe and so what I do with caps of my pens with this feature is use them as I would a hair-clip, to hold up my fringe.

How one can take a bit of rubbed off flesh from the inside of your cheek and use it to figure out where some of your ancestors came from is also quite extraordinary. And yet it is rote work for geneticists.

Imagine the process of genotyping some individuals, specifying markers in their genetic material, coding and storing their data, to the point when the data is clustered by STRUCTURE.

Take a swab of say ‘Mary’s body matter’. Through the physical processes of sampling and sorting, *that* matter will physically lose some of its possible uses and correlatively some of the possible meanings that it originally could have had as, say, ‘the inside of Mary’s cheek’ [of some use to Mary, some other to Mary’s dentist, another to her digestive system, etc.] and retain the ones it has as ‘Mary’s epithelial cells’ on the 4^{th} stick in this sampling situation. This sorting of meanings will continue with each consecutive sorting of the swab -say as ‘50^{th} saliva swab’, ‘60^{th} DNA sample’, ‘genotype of the i^{th} individual at the l^{th} locus’, ‘D13s119 (1310)’, etc, for each of the swabs collected, until the material on these swabs is meaningful as ‘random draws’: marks which are read and sorted by the algorithm of STRUCTURE –that does so according to a model developed by the biostatisticians to capture and represent ‘genetic’ structure.

What happens at the same time as processing *materials* through various arrays and vessels and chemical solvents, is that we (geneticists) essentially and sequentially sift through the possible *uses of* the materials sampled until that material (or rather a form of that material) is *usable* as data for STRUCTURE’s algorithm; i.e. marks that occur with particular frequencies that the algorithm can sort through.

This dissertation is concerned with concepts of race and my development of the idea of found science is confined to founding ‘ordinary‘ race concepts into biomedical scientific contexts here. But the ideas and structure seem appropriate –and helpful– in thinking about certain objects as well. And in this case, thinking of the objects this way helps see how the concepts get founded in the new scientific context. I described my rote use of a pen-cap as a hair-clip as resulting in a founded ‘pen-cap’: i.e. a pen-cap founded in the context of use of a hair-clip, and used as a found ‘hair-clip’. Similarly, through each step in the process of sorting individual organisms’ DNA swabs and correlatively to the various uses they are put to, each progressively sheds possible meanings (meanings each would have in different contexts) by virtue of the thing, the swab, moving to different contexts. The body matter is a found ‘DNA sample’, a found ‘genotype’, a found ‘genotype at a particular locus’, a found ‘datum’; and correlatively gets founded in different context of use.

So here is why I think we could call STRUCTURE a founding tool for genetics: Developed according to the precepts and interests of biostatistics STRUCTURE is designed to "pick up" features in the data defined as what should make 'genetic' sense. This is not a necessary 'meaning' the samples have at this stage –or at any stage of the process; I might have taken a photo of the magnified molecules and hung it on my bedroom wall thus finding these samples as 'décor'. STRUCTURE does no such thing.

STRUCTURE "articulates" the structure of carefully collected data in terms of a predefined -genetic- mathematical model. And STRUCTURE helps perform this bit of articulation in a process which started with say {'Mary's body matter', 'Troy's body matter', 'Julie's body matter', etc…} and ends up with 'K genetic populations' using minimal information derived from the data's original context.

It seems astonishing. But one shouldn't get too invested in a tool. Of course a crane can pick up a shark's form but one has to operate the crane and make sure it is the shark that is in the crane's grasp. Similarly in this context: all the 'non-genetic' information is sequentially filtered through via checks on what is relevant and what is not for this 'genetic' usage, at every step of the process, until the information relevant to genetics remains in the form of data readable by STRUCTURE.

So, using this clustering algorithm, within this particular bioinformatics context, finds data as of genetic interest: it abstracts data from their original ("subjective") contexts, [rendering them (saliva swabs) meaningless in those (drooling) contexts], and at the same time structures the data according to a model already formulated according to the precepts of genetics[103] as one that renders the data articulate in the terms of genetics[104]. So this is a founding tool: it can "pick up" features of interest to genetics and install them in a genetics context.

---

[103] Pritchard cites his collaboration with Noah Rosenberg in Pritchard and Rosenberg (1999) in developing a model predating this one.

[104] The process I described (for reasons of simplification) as one of genotyping leaves a paper trail; samples are collected and moved around following protocols, filling out forms, etc. I mention this because emphasis has been placed on these inscriptions as of interest by Latour and Woolgar (1986). There is no question that these are important. But these inscriptions are only as important as traces in a murder mystery are: they are not the killer nor the dead. What is important about these forms is that they *specify* what about this sample is –or if you want, is taken to be– relevant for this context of use. What *is* decided as relevant is not simply

Note again the limitations of such a process.

The fact that STRUCTURE can articulate data into genetic clusters and found them in a genetics context does not ensure that the clusters obtained will be biologically meaningful; nor does it ensure that the clusters will be useful for genetics. Remember that found science is a story about aptness rather than merit. (Merit is discussed in the next episode once what we call "science" is set). Whether the clusters are biologically meaningful is contingent on a number of factors, like the number of individuals sampled, the number of loci used, the amount of admixture in place and the extent of allele-frequency differences in the populations[105]. The biological meaningfulness of the clusters depends on the "appropriateness" of the samples. And the usefulness of the clusters depends on further interests.

3.3.1.1.3 How Could STRUCTURE Make Genetic Sense of Ordinary 'Populations'? I.e. Why is it Usable as a Tool for Founding Ordinary Race Concepts?

Recall the two projected uses hoped of STRUCTURE (3.3.1.1.2):

a. STRUCTURE can be used to check if "subjective" classes *naturally* make sense in genetic terms.

b. STRUCTURE can be used to confirm that "subjective" classes are apt for answering questions of interest to genetics.

Thinking of this tool as sorting samples until these shed ordinary meanings and become articulate in genetic terms justifies the claim that what population clusters are obtained using STRUCTURE are of a form "appropriate to genetics".

But projected uses a. and b. above have "subjective classes" in their target: using STRUCTURE should confirm that ordinary *population* concepts are appropriate for answering questions of genetic interest.

---

a matter of inscription but rather of the possibility of such an "inscription", and one that makes sense in a genetics context. Of course people get trained to see tracks and infer being; but this does not mean that the tracks are what should fascinate us.

[105] Cf. Pritchard et al (2000), 955.

But how is that possible?

STRUCTURE can infer the structure of *data* in genetic terms. It can structure data into genetically articulate populations. But who can infer the structure of *populations* in genetic terms? On what logical basis can these researchers expect "the method developed here" to be used in these ways?

3.3.1.4 A RECIPE for Founding Ordinary Population/Race Concepts via STRUCTURE

STRUCTURE operates on and "picks up" data; not populations. Sorting populations involves another sorting process, one that happens outside STRUCTURE's model. Below is a recipe that would justify the use that Pritchard et al (2000) hope to make of STRUCTURE.

The consistency of subjective population concepts with genetic information could be established by using STRUCTURE as follows.

RECIPE (Rules of Embedding Commonly Interesting Population Concepts) for Genetics

0. Sort genetic data into K populations using articulation tool STRUCTURE.
1. Let the populations obtained in (0) be a "genetic check list" of K genetic populations; a standard of genetic sensibility. –This operates on the assumption that STRUCTURE is a finding tool for genetics; it can pick up genetically interesting structure.
2. Check ordinary ("subjective") population concepts against the check-list obtained in (1). – A map for "checking" needs to be defined along with error margins.
3. If the ordinary population concepts "check out" as approximately enough picking out the same K populations call them "natural in genetic terms". –This makes the populations found 'genetic populations' in my terms. Let the 'genetic' founding of an ordinary concept be the image of the class picked out by this ordinary concept under the designated checking map. This demonstrates that STRUCTURE can be used as a founding tool to bring the ordinary concepts into the space of genetics. An ordinary population concept thus articulated would be a founded concept; founded in the space of genetics.
4. Call the ordinary, genetically articulated classes found in (3) "appropriate for answering the questions of genetic interest" –This amounts to deeming these categories 'genetically relevant'; they can speak to questions of genetic interest. This demonstrates that STRUCTURE can be used as a founding tool for a context of interests of genetics.

> Note that founding a concept is a matter of founding it in both the spaces of genetics and in the context of interests of genetics. Until found genetic classes are used to answer a genetic question, these concepts may not be founded in a context of genetic interests. Being "articulate" in genetic terms is necessary, not sufficient, for "speaking to" a question of genetic interest. Or in my terms, finding an ordinary class in the context of genetics is not sufficient for founding it there as genetic; one can be articulate and silent but not answer a question lacking articulation.

This is a recipe, what Weber calls an "ideal-typical" rule[106], for how these researchers could be using STRUCTURE to sort ordinary populations according to their genetic appropriateness. It gives a rule –a norm- for when ordinary population concepts could be imported to a context of interests and uses particular to genetics. It doesn't ensure that the ordinary populations will be used to answer questions of interest to genetics. But it gives you that option. Like checking the color of an object against a color inventory (step 2) allows us to call it 'red' or not (step 3) and so gives us the option of referring to this object in answering questions about 'red' things (step 4), checking population concepts against this genetically informed population inventory allows us to call them 'natural in genetic terms' and look to this select class of ordinary population concepts if we care to use ordinary population concepts to answer questions of interest to genetics[107].

3.3.1.5 An Incomplete RECIPE: Missing Detail and Steps

This is an ideal norm; it is not proposed as an accurate description of what is going on. What actual norms constrain these scientists' practice is a matter of empirical research. But I propose this is a way of describing what logic their practice follows. Further, like any recipe it leaves a lot of terms undefined: importantly, it does not define what an appropriate checking map and error terms would be.

---

[106] Weber (1978 [1904])

[107] Wittgenstein defines a 'language-game' by using a similar example. Wittgenstein (2001 [1953]). I resist the terminology because 1. "game" is a misleading name for this process; it underestimates the seriousness of such 'games' and their usefulness in structuring what Wittgenstein might call particular "forms of life" and such games are not entered into lightly by the "players", 2. "language" deflects from the importance of what concepts go into the game and the ontological import of playing not only with words but with concepts. So this is a poor term for this context; unless one is prepared to understand the term "language-game" ironically.

First, checking against a list can be done in different ways and second, deciding when something "checks out" is a matter of degree –it involves what we call errors. For example, in checking the color of a thing against our inventory of 'red', we might further specify 1. check the color by placing the thing next to the inventory making sure the lighting on the inventory and thing is of the same intensity and hue judging whether the transition in looking from one to the other looks 2. from any distance/from 10 cm away/from 3m away/etc. smooth. Alternatively we might require 1. a color expert, someone who has a correct knowledge of colors, retains an immaculate memory of perceived colors and has an ability to match the two, to do the matching of 'color' to 'red'. And again the expert's proclamation can be quantified by 2. the degrees of his belief in it/its correspondence to past records of the object's 'color'/etc. Or we might check a thing's color by 1. dropping it on a color inventory and calling it 'red' if it 2. falls squarely within/touches/touches only/etc. the 'red' entry. Etc. Ranking each of these checking methods and the degree of specificity of their claims in terms of their appropriateness for the purpose at hand is another enterprise; one which science is in the business of.

Further, were this recipe to be followed it would be found lacking in yet another way: There are steps missing. For example, having correctly matched some thing to 'red' what do we do with it? Why were we interested in picking out its 'color' to begin with? An interest in waving the thing in front of a bull in a bullfight will demand that that the thing, found to be red, should be appropriate for waving in front of a bull in a bullfight. Given our prior knowledge of bullfights this will exclude a (red) strawberry, a (red) shoe, a (red) apple, etc. Likewise, these instructions do not say what we are to do with the populations that are appropriately checked against our standards, found to be 'genetic', or more precisely genetically articulate, and checked into (and possibly later out of) the context of genetics. RECIPE gives mere entry rules for an ordinary population to enter the context of interest and use of genetics but it says nothing about what use is to be made of and what meaning is to be founded for an ordinary population in a genetic context.

Further instructions guided by further demands will further sort through the ordinary categories which are deemed genetically articulate and decide which ones have something to say to the questions at hand (Step 5, 6, 7…etc)[108].

So it must be granted that I have given an incomplete recipe for sorting populations according to their genetic aptness. And this is because I have not specified 1. a method for checking populations against the genetic check list and I have not specified 2. with what degree of error the proclamation of "checking out" should be made, and 3. For what purpose the checked out materials will be used. But *how* the recipe can be followed will become clear in the next two sections, where I argue it *is* –roughly– followed to articulate human populations in genetic terms because these may be of biomedical/biological interest.

3.3.2 Ordinary 'Populations' Being Founded in Genetics

STRUCTURE was developed to sort through genetic variation, find K clusters of genetically similar populations and then see which of these, if any, match our subjective categories –thus founding these subjective categories into contexts of genetic enquiry.

But what are the "questions of interest" that STRUCTURE is used to answer?

STRUCTURE is used to answer questions of interest to human population genomics, such as: What is the genetic structure of *human* populations?

I examine two studies which used STRUCTURE to determine the genetic structure of human populations. In both cases I argue that: a) STRUCTURE is used according to a RECIPE to check if common population categories are relevant to the context of genetics, b) prior interests are used to sort through categories that might be genetically relevant –adding steps to RECIPE- and c) these interests do not originate in the context of genetics practice and they are not logically mandated by genetic data.

Recall, these would be the ingredients of found science:

[108] That being said, getting stamped with a "genetic" label may in fact get you a long way away, in this day and age…

A. Ordinary, non-genetic concept

B. Two types of action

1. Finding

2. Founding

C. Two types of context stamped by science,

(a) a Spatial Context and

(b) a Context of Interests

D. Tools for performing an action, according to context:

1. Finding Tools for C.(a), for C.(b) usable as

2. Founding Tools for C.(b), for C.(a)

By specifying a tool, STRUCTURE, and describing how its use fits the context of genetics I have simplified the picture radically but sorted B,C,D. What are the ordinary concepts in question?

These are the ordinary concepts discussed in the context of the two studies of Rosenberg et al (2002) and Tang et al (2005):

A1. Ordinary, non-genetic concept in Rosenberg et al. (2002): 'Self-identified ancestry'

A2. Ordinary, non-genetic concept in Tang et al (2005): 'Self-identified (OMB) race/ethnicity' or 'SIRE' for short.

And the conclusions of found science in this case are the following two claims:

1. 'Self-identified ancestry' is founded in a genetic context by Rosenberg et al. 2002: it is found 'genetic ancestry'
2. 'Self-identified (OMB) race/ethnicity' is founded in a genetic context by Tang et al 2005: it is found 'self-identified ancestry', founded by Rosenberg et al as 'genetic ancestry'.

'Race' is getting a legitimate place in genetics research by being sequentially found and founded in the contexts of space and interests of genetics.

3.3.2.1 'Self-Identified Ancestry' is Found 'Genetic Ancestry'

The study of Rosenberg and colleagues 2002 reports the "Genetic Structure of Human Populations"[109]. The authors report a fit between *five* major geographic regions and clusters obtained using STRUCTURE: 1. America, 2. East Asia, 3. Pacific Islands, 4. Africa and 5. Eurasia.

The study is widely cited in the literature as it corroborated estimates posited previously[110] that the majority of human genetic variation occurs within populations rather than between them[111]. At the same time, the study that confirmed a statistic widely cited as an indication of overall human similarity did, I argue, make a step towards founding common race/ethnicity categories in genetics. How?

By founding the notion of 'self-identified ancestry' in the context of genetics.

3.3.2.1.1 Using STRUCTURE by a RECIPE: STEPS 0, 1= Sorting Data, Getting a Genetic Check-list

In their 2002 paper Noah Rosenberg and colleagues[112] used software STRUCTURE, to sort genotypes at 377 autosomal microsattelite loci of 1056 individuals from 52 populations across the world[113]. STRUCTURE was supposed to be used to pick out genetic structure in this sample so the authors titled their study: "*Genetic* Structure of Human Populations" [emphasis added]. This would be in line with STRUCTURE's use as a founding tool for genetics.

Indeed Rosenberg and colleagues echo the concerns of Pritchard et al. (2000) regarding the correspondence of ordinary population classes to "underlying genetic relationships".

---

[109] Rosenberg et al. (2002)
[110] Cf. Lewontin (1972), Bowcock et al. (1994), Barbujani et al. (1997). The last one of these also used microsatellite markers to infer population structure. They did not use program *structure* but rather nonparametric statistical methods for analysis of molecular variance (AMOVA).
[111] Rosenberg et al (2002), p 2381. The fraction of within population genetic variation estimated here was higher than previous estimates; a result which the researchers justify by hypothesizing "differences in sampling schemes". Of course this widely cited claim is not as indicative of the meaninglessness of racial difference as many assume; the fact that most genetic variation is within races does not imply there is no genetic variation between ordinary races nor that there is no interesting variation between these groups.
[112] Co-authors included the developer of STRUCTURE, Jonathan Pritchard.
[113] See Appendix 1 for definitions of some genetics terms including "genetic marker".

> Most studies of human variation begin by sampling from predefined "populations." These populations are usually defined on the basis of culture or geography and might not reflect underlying genetic relationships (2381).

STRUCTURE can succeed in abstracting data from its ordinary context and picking out the data's genetic structure:

> Despite small among-population variance components and the rarity of "private" alleles, analysis of multilocus genotypes allows inference of genetic ancestry without relying on information about sampling locations of individuals (2382).

3.3.2.1.1.1 Using STRUCTURE as a Founding Tool

The genetic data used in this case was obtained from a genetic data bank of the Foundation Jean Dausset in Paris (CEPH)[114]. This data derived from 1064 cultured lymphoblastoid cell lines (LCLs) obtained from 1051 individuals from 51 populations across the world. The DNA of each of those cell lines has been typed with 404 microsatellite markers[115]. The data is anonymous but labeled by the information (sex(i), population(i), geographic location(i)) for each individual, i[116].

This sampling is already selective. It assumes a selection of "appropriate" individuals for which these sets of characteristics are well-defined and a selection of "appropriate" characteristics for each of the individuals sampled. What guides the selection of this information as "appropriate" are entry rules specific to the context of interests and use of genetics. So, this is already a founded notion of 'individual'; shrunk to three 'nongenetic' but genetically relevant markers. But this is not the articulation of 'individual' data on which STRUCTURE operates.

STRUCTURE operates on the mathematical articulations of the genetic data which these 'individuals'' labels label. Sequential sorting of the possible meanings of say 'swab of i^{th} individual', 'LCL(i)', 'DNA(i)', 'genotype of the i^{th} individual at the l^{th} locus', etc. find as ordinary and found as genetically interesting the data obtained; until it makes sense in terms of the mathematical symbols that STRUCTURE's algorithm processes.

---

[114] http://www.cephb.fr/HGDP-CEPH-Panel/ (Last accessed on 13/4/07)
[115] See Appendix 1 for definitions of some genetics terms.
[116] For a table with the populations characteristics see: http://www.cephb.fr/HGDP-CEPH-Panel/HGDP-CEPH_Table1-1.html (Last accessed on 13/4/07)

The application of STRUCTURE is described as follows:

> We applied a model-based clustering algorithm that, loosely speaking, identifies subgroups that have distinctive allele frequencies. This procedure, implemented in the computer program *structure*, places individuals into K clusters, where K is chosen in advance but can be varied across independent runs of the algorithm. Individuals can have membership in multiple clusters, with membership coefficients summing to 1 across clusters (2382).

This description is loose. The researchers claim that the procedure implemented by STRUCTURE "places individuals into K clusters" and that "individuals can have membership in multiple clusters, with membership coefficients summing to 1 across clusters" (b). But what are these 'individuals' referred to? Are they 3-ples of (sex(i), population(i), geographic location(i)) or say '$\sum_{j=1}^{K} a_{ij}^{n} C_{j}^{n}$' some linear combination of their cluster membership coefficients inferred at an n^{th} run of STRUCTURE ,etc.?

The only stuff that STRUCTURE actually places into clusters is mathematical symbols. But this is not an answer that troubles these scientists. And this is because although each of these representations are taken to be genetic articulations of a common notion like an 'individual' they are articulated, articulate and speak to different stages in the practice of genotyping: they are founded notions and they make sense in the context of genetics. It is only an outsider who might be confused. And this is because they lack the story behind the names. This is a founded notion of a founded notion of an individual that researchers would be operating with; not any ordinary one(s). [Of course this doesn't make the authors' assumptions correct.]

The clusters into which data is sorted by STRUCTURE are identified as "subgroups" that have "distinctive allele frequencies" (2382). These 'subgroups' are again clusters of data. What would logically ground an inference from data subgroups to found 'population' subgroups is summed up in STRUCTURE's design specifications and the warranty it comes with in Pritchard et al (2000) but is also based on background theory and practices. These clusters are articulate in genetic terms because they have been obtained in accordance with a genetic model –another founding tool- which founds and renders mathematically articulate the notion of 'genetic population subgroups'.

> Despite small among-population variance components and the rarity of "private" alleles, analysis of multilocus genotypes allows inference of genetic ancestry without relying on information about sampling locations of individuals (2382).

So this claim relies on a lot of background theory but it is genetics theory.

But how are common race concepts being founded in the context of genetics here?

I answer in two steps.

3.3.2.1.1.2 STRUCTURE as a Founding Tool

The use of STRUCTURE posited in Pritchard et al (2000) was that of confirming that certain "subjective" classes can be used to answer questions of interest to genetics. This type of use is of interest too to Rosenberg et al. (2002) in the case of humans, and in particular in studying the evolutionary history of modern humans:

> Because knowledge about genetic structure of modern human populations can aid in inference of human evolutionary history, we used the HGDP-CEPH Human Genome Diversity Cell Line Panel to test the correspondence of predefined groups with those inferred from individual multilocus genotypes (2381). [Emphasis added]

What are the pre-defined groups supposed to be and what does STRUCTURE relate these to?

Two notions are important here. First we read that what this genetic method allows is inference of "genetic ancestry":

> Despite small among-population variance components and the rarity of 'private' alleles, analysis of multilocus genotypes allows inference of genetic ancestry without relying on information about sampling locations of individuals" (2382). [Emphasis added]

This is the genetically sensible concept: 'genetic ancestry'. What is the ordinary notion related to 'genetic ancestry' here?

We read in the conclusions of the study that

> General agreement of genetic and predefined populations suggests that self-reported ancestry can facilitate assessments of epidemiological risks but does not obviate the need to use genetic information in genetic association studies (2381). [Emphasis added.]

I said this answer has two steps; for now let us stick with the first step which is the ordinary notion of 'self-reported ancestry' or, I think equivalently, 'self-identified ancestry'.

How are these two concepts found to be in correspondence?

3.3.2.1.1.3 Using STRUCTURE by RECIPE

A RECIPE is used to examine the relevance of 'self-reported ancestry' for genetics. What is 'self-reported ancestry' in this context? This is captured as 'population(i)' for each individual's sample. [Information on the "regional affiliation" of this population, captured as location(i) for each individual's sample matters later.]

As suggested in the previous section by feeding *this* same genetic data (*thus* labeled) into STRUCTURE 'genetic population subgroups' and 'predefined populations' are found to be in "general agreement". This involves using statistical methods to compare the 'genetic population subgroups/cluster' membership coefficients obtained for individuals in the same pre-identified groups to each other, in different runs of STRUCTURE. Figure 5 below represents how STRUCTURE clustered data at different values for K –specified on the left.

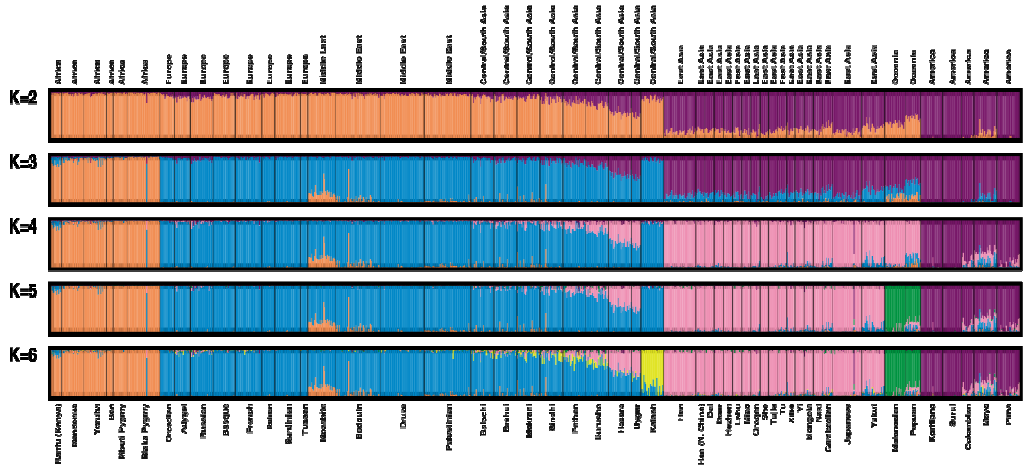


Fig. 1. Estimated population structure. Each individual is represented by a thin vertical line, which is partitioned into *K* colored segments that represent the individual's estimated membership fractions in *K* clusters. Black lines separate individuals of different populations. Populations are labeled below the figure, with their regional affiliations above it. Ten *structure* runs at each *K* produced nearly identical individual membership coefficients, having pairwise similarity coefficients above 0.97, with the exceptions of comparisons involving four runs at $K = 3$ that separated East Asia instead of Eurasia, and one run at $K = 6$ that separated Karitiana instead of Kalash. The figure shown for a given *K* is based on the highest probability run at that *K*.

Figure 5: Results for human genetic structure for K-2 through K-6.

We read in the caption to this graph:

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

> pairwise similarity coefficients above 0.97 with the exceptions of comparisons involving four runs at K=3 that separated East Asia instead of Eurasia, and one run at K=6 that separated Karitiana instead of Kalash. The figure shown for a given K is based on the highest probability run at that K (2382).

RECIPE (Rules for Embedding Common Population Concepts) for genetics would read as follows if applied to 'self-identified ancestry' classes:

0. *Sort genetic data into K populations using articulation tool STRUCTURE.*
1. *Let the populations obtained in (0) be a "genetic check list" of K genetic populations.*

   Here we don't care about the structure of this genetic material simpliciter; rather we care about the structure of this genetic material as anchored by individuals. So, we must look at how genetic clustering manifests on the level of individuals. As individuals' genetic data is sorted into K populations every individuals' sample is fully partitioned into K clusters and so for every individual sampled membership coefficients in the K clusters will sum to 1. The distinctive set of membership coefficients in each of the K clusters that each individual has is compared to that of other individuals in the sample, (and that of the same individual in another run of structure for the same K) and the coefficients matched using statistical comparison methods. And thus it is the membership coefficients for each individual to every cluster that are obtained as of genetic sensibility and particular structure in what these coefficients are that is identified. This still results in roughly K populations.
2. *Check 'self-identified ancestry' classes against the check list obtained in (1).*

   A map for checking is defined.

   'Predefined populations' are obtained from entries in the 3ple that labels each individual's genetic sample: population(i), in the i^{th} individual's sample, is recorded for every i. Then for every individuals k and j in the same cluster in (1) it is examined whether their predefined labels 'population(k)', 'population(j)' match.

   Authors note that "In the worldwide sample, individuals from the same predefined population nearly always shared similar membership coefficients in inferred clusters"

(2382). This is an inference which legitimates the relevance of predefined population labels, or 'self-identified ancestry' specifications, for genetic interests.

Further the degree of the match, or how many loci were needed to produce these clusters mattered. The researchers were interested in knowing how much detail would be enough detail –thus specifying margins of tolerable error for such a process. For example,

> The amount of among-group variation affects the number of loci required to produce clusters similar to those obtained with the full data. …Fewer loci would probably suffice for larger samples; conversely, accuracy decreased considerably when only half the sample was used. The number of loci required would also decrease if extremely informative markers, such as those with particularly high heterozygosity, were genotyped (2384).

3. *If the ordinary population classes "check out" call them "natural in genetic terms" –This makes the populations found 'genetic populations' in my terms.*

   This is a correspondence established with a concept formulated in a context of space of genetics: so this step amounts to founding 'self-identified ancestry' in this context. This demonstrates that STRUCTURE is used as a founding tool for a context of space of genetics.

   Let the 'genetic' founding of an ordinary class be the image of this ordinary class under the designated checking map.

   There is no such inference made; the particular loci used and the particular genetic structure obtained are not investigated further here. The proclamation made is of there being a correspondence between the ordinary classes and genetic ones with no specification of what the image of the ordinary classes in genetic terms would be under this checking map.

4. *Call the ordinary, genetically articulated classes found in (3) "appropriate for answering the questions of genetic interest" –This amounts to deeming these categories 'genetically relevant'; they can speak to questions of genetic interest. Note however that until found genetic classes are used to answer a genetic question, these classes remain unfounded in a context of genetic interests. This demonstrates that STRUCTURE can be used as a founding tool for a context of interests of genetics.*

   This is an inference drawn by the authors. The authors draw a conclusion about what 'self-reported ancestry' can do: it can facilitate assessments of epidemiological risks.

> We have found that predefined labels were highly informative about membership in genetic clusters, even for intermediate populations, in which most individuals had similar membership coefficients across clusters. Sizable variation in ancestry within predefined populations was detected only rarely, such as among geographically proximate Middle Eastern groups (2384).

They continue to say:

> Thus, for many applications in epidemiology, as well as for assessing individual disease risks, self-reported population ancestry likely provides a suitable proxy for genetic ancestry (2384).

They justify this claim by considerations that may be relevant in such a context, when dealing for example with human patients:

> Self-reported ancestry can be obtained less intrusively than genetic ancestry, and if self-reported ancestry subdivides a genetic cluster into multiple groups, it may provide useful information about unknown environmental risk factors. One exception to these general comments may arise in recently admixed populations, in which genetic ancestry varies substantially among individuals; this variation might correlate with risk as a result of genetic or cultural factors (2384).

The claim made manifestly is that 'self-identified ancestry' is likely a good proxy for 'genetic ancestry'. This has been inferred on the basis of a particular embedding of the notion of self-identified ancestry in the context of genetics: by taking individuals' self-identified ancestry to stand for the label 'population(x)', that labels their data, stored in CEPH. And then matching genetic variation to this particular founded notion of 'self-identified ancestry' is assumed to confirm that the ordinary notion can be of relevance, and a found 'genetic ancestry', where questions are of genetic interest.

Further, this conclusion would be manifestly in agreement with Michael Root's (2003) argument that ancestry rather than race is a better proxy for medically interesting genetic variation. However, in the process of justifying why 'self-reported ancestry' may be relevant for epidemiological work Rosenberg et al cite research done on the significance of race/ethnicity labels for biomedical and genetic epidemiological work by Risch et al (2002), which is the work that Daar and Singer used to further their claims for the usefulness of biorace for medical research (cf. Section 2.2).

3.3.2.1.1.4 'Self-identified ancestry', at K=5 v. 'race'

Rosenberg et al (2002) speak of 'self-reported ancestry' rather vaguely in their article, but there is a specific fineness of grain at which 'self-reported ancestry' seems to matter a bit more, and this happens to be a fineness of grain that could correspond to a biorace notion.

Rosenberg and colleagues report the fit found between *five* major geographic regions and *five* of the clusters that occur at K=6 in their abstract. The regions clusters correspond to are: 1. America, 2. East Asia, 3. Pacific Islands, 4. Africa and 5. Eurasia. This is how they describe their results in the article's body:

> At K=5, clusters corresponded largely to major geographic regions. However, the next cluster at K=6 did not match a major region but consisted largely of individuals of the isolated Kalash group, who speak an Indo-European language and live in northwest Pakistan (2381). [Emphasis added]

Something it seems goes contrary to expectation in moving from the results at K=5 to those at K=6. The authors explain why the manifest genetic distinctiveness of the Kalash was most likely due to a limitation of STRUCTURE's model:

> Because genetic drift occurs rapidly in small populations, particularly in those that are also isolated, these groups quickly accumulate distinctive allele frequencies. Thus *structure* efficiently detects isolated and relatively homogeneous groups, even if the times since their divergences or exchanges with other groups are short. This phenomenon may explain the inferred distinctiveness of groups with low heterozygosity such as Lahu and American groups, and those that are small and isolated, such as Kalash (2383-2384).

So, though the Kalash come out as a distinctive population at K=6 this is more likely because they have been isolated and small than because they are genetically distinctive.

It seems reasonable to say that K=5 is selected and reported in the abstract because it is the maximal, finest grain at which STRUCTURE did not make a mistake. Further, there is reason to think these clusters are of interest to evolutionary theory as they correspond to major geographic boundaries.

> Because sampling was population based, the sample likely produced clusters that were more distinct than would have been found in a sample with random worldwide representation. However, world-level boundaries between major clusters mostly corresponded to major physical barriers (oceans, Himalayas, Sahara) (2383-2384).

And they conclude that

> The challenge to genetic studies of human history is to use the small amount of genetic differentiation among populations to infer the history of human migrations…. Patterns of modern human population structure discussed here can be used to guide construction of historical models of migration and admixture that will be useful in inferential studies of human genetic history (2384).

Rosenberg et al. use a prior assumption about what geographical regions matter to sort through the clusters of genetic variation picked out by STRUCTURE and to decide which ones are worth reporting about in their abstract. They would most probably argue that this prior assumption is our evolutionary story (our theory) of what should count as ancient geographical ancestral groups[117].

This doesn't mask the fact, (it might even be historically *due to the fact*), that these geographical regions match continental origins of what we commonly take to be five human races[118].

A second look at figure 5 helps make this point. Though samples came from many different populations (labeled on the bottom) there seems to be significant homogeneity in individuals' cluster membership coefficients in samples coming from the same major geographical regions (labeled on the top). There is no explicit mention of race in the article. But noting what these major regions are lends some basis to infer that a race notion, specifying groups of people on the basis of HLC2 and HLC3 –criteria specifying ancestral ties and correlative geographical continental origins for these lineages– may be of relevance in this context: if one counts the colors for K=5 and looks up to the "regional affiliations" of the populations on the top of the figure on sees that they more-or-less map onto the continental origins of common races.

It seems that 'self-identified ancestry' groups, found to approximate 'genetic ancestry' groups (so satisfying HLC2) who also originated in the K=5 regions (so satisfying HLC3) would be conceptually close to races –or better self-identified races. Though phenotypic features were not recorded here one wonders whether this would be once more as in the case of OMB race/ethnicity (cf. Appendix 3), a criterion dropped from manifest mention.

---

[117] Cf. Hardimon's communication with Rosenberg.
[118] Cf. Müller-Wille and Rheinberger (2007). This is unsurprising if we accept that the notion of 'race' could already be founded in notions of genetic heredity.

Explicitly, Rosenberg et al (2002) do not attempt to found races in the context of genetics; explicitly this is an attempt to check if 'self-identified ancestry' makes genetic sense; but the conceptual affinity between 'race', or 'self-identified race', and 'self-identified ancestry' at K=5 begs further scrutiny.

3.3.2.1.1.5 Missing Steps in RECIPE

What conclusion can be drawn though is that, as expected, the RECIPE is incomplete. It takes input from interests not specific to genetics before a use for these found genetic ancestries can be decided.

There is explicit gesturing in Rosenberg et al. (2002) towards the *medical* significance of ancient geographic ancestry. Out of the K=2 through K=7 clusters that are all picked up as potentially genetically relevant it is the fineness of grain that maps onto OMB race/ethnicity that is held –or sustained– as of interest to biomedical genetic research. Though they do not engage in a critical argument Rosenberg et al (2002) claim that "self-reported ancestry can facilitate assessments of epidemiological risks"[119].

Commenting on this article John Hardy of the NIH predicts that "genetic techniques such as this will facilitate the definition of *ethnic groups* based on genomic variation and enable scientists to test the hypothesis that diseases have divergent clinical features between these groups"[120]. The hope is that some of the alleles that cluster into different frequencies for these different groups will be causally related to disease susceptibility or to response to treatment. These remarks suggest that the context of interesting questions that STRUCTURE can be used to answer can be stretched further.

So, we took a tool developed to articulate structure in genetic terms to articulate the structure of human populations in genetic terms and the suggestion is that the same tool can be

[119] Rosenberg et al. (2002), 2381. Still, it is unclear that the 3-5% of genetic variability occurring between these populations will have something to do with disease. See Keita et al. (2004) on competing hypotheses regarding the medical significance of this genetic variability. This article appears in a special supplement of Nature Reviews Genetics on race and genetics, published in 2004, and used to install these questions in a discursive context of genetics.
[120] Hardy (2003) 739

used to explore human populations' *medically* interesting structure in genetic terms. This context of interests includes questions of interest to *both* human population genomics and *epidemiology*: Do different race and ethnicity groups capture medically interesting genetic structure?

That self-reported ancestry can facilitate assessments of epidemiological risks relies on research performed on particular populations tagged by 'OMB race/ethnicity' which is (cf. Appendix 3) a founded race notion.

It is with attention to this founded notion that Rosenberg and colleagues' claims are made. Considerations that may be relevant in such a context, when dealing for example with human patients, do regard the invasiveness (and costliness) of such a procedure:

> Self-reported ancestry can be obtained less intrusively than genetic ancestry, and if self-reported ancestry subdivides a genetic cluster into multiple groups, it may provide useful information about unknown environmental risk factors. One exception to these general comments may arise in recently admixed populations, in which genetic ancestry varies substantially among individuals; this variation might correlate with risk as a result of genetic or cultural factors (2384).

This indicates a concern for sustaining a relationship between the ontologies of the genetics and epidemiological context. But this would be a relationship that needs to be nourished and fortified with further research.

A particular study in genetics such as this one, of Rosenberg et al (2002), should not constitute epstemologically sufficient grounds for importing 'self-identified ancestry' in the context of epidemiology. This study indicates that the founded concept can function much like OMB race/ethnicity categories are found to function in epidemiological contexts. However OMB race/ethnicity, as used in epidemiology is a founded (founded) race notion. It is –or should be– only via particular processes endorsed and developed to attribute epidemiological meaningfulness to non-epidemiological, genetic constructs that such categories should be handled in a target epidemiological context. Just because the categories come from a discipline like genetics with founding tools and concepts that are more honed into the material substratum of phenomena, doesn't make them relevant to an epidemiological context.

Interdisciplinary relationships stand to benefit all involved; but they need to be recognized as different kinds of work than the work already happening within each discipline for it to develop, find and found, its ontology.

3.3.2.2 'Self-Identified Race/Ethnicity' is Found 'Self-Identified Ancestry'

Rosenberg et al link up 'genetic ancestry' with 'race' via the notion of 'self-identified ancestry'. And they don't do this explicitly –rather only via quoting results in the biomedical literature obtained under self-identified race/ethnicity categories.

But, forget nested meanings! We don't need to look too far for an explicit genetic interest in race. In a paper titled "Genetic structure, Self-Identified Race/Ethnicity, and Confounding in Case-Control Association Studies" Hua Tang and colleagues at Stanford University examine whether self-identified race/ethnicity approximates genetic structure in the USA. They answer in the positive.

In this study subjective population classes were explicitly defined in terms of self-identified race/ethnicity. These results were obtained using STRUCTURE but also using genetic distance analysis, i.e. measuring the coancestry coefficient among groups identified within the context of the study as of possibly genetic relation. For reasons of simplicity and consistency with the previous study I focus on the results obtained using STRUCTURE.

Individuals self-identified as belonging to four major 'OMB race/ethnicity' groups: white, African American, East Asian, *and Hispanic* –the OMB ethnicity. 326 microsatellite markers were taken; subjects came from 15 different locales within the US and Taiwan. Using STRUCTURE implied that: of 3,636 subjects of varying race/ ethnicity, only 5 (0.14%) belonged to a genetic cluster different from their Self-Identified Race/Ethnicity (or SIRE) –where SIRE is the OMB classification.

This result quoted in the article's abstract suggests a link between SIRE and genetic variation:

> [A]ncient geographic ancestry, which is highly correlated with self-identified race/ethnicity —as opposed to current residence— is the major determinant of genetic structure in the U.S. population (268).

Once more it is not 'race' but rather 'ancient geographic ancestry' which is reported as the major determinant of genetic structure in the U.S. But ancient geographic ancestry is now found to be "highly correlated" with self-identified race/ethnicity.

STRUCTURE was used following a RECIPE to relate 'self-identified ancestry' and 'genetic ancestry' in Rosenberg et al (2002). As we saw, Rosenberg et al. (2002) selected K=5 genetic clusters picked up by STRUCTURE as determined by 'genetic ancestry' corresponding to major geographic, continental regions. Tang et al. (2005) report K=4 classes picked up by STRUCTURE as likewise matching 'ancient geographical ancestry'. But this claim doesn't rest on evolutionary theory but is rather based on their correlative use of the method to check the correspondence of clusters to more recent geographical origins of the individuals sampled, such as place of residence. And clusters corresponded more tightly to the ancient geographical origin of the individuals sampled than to their place of current residence.

3.3.2.2.1 Founded Concepts

The authors seem alert to the dubious nature of the groups thus identified. They note that:

> Another major point of discussion has been the correspondence between genetic clusters and commonly used racial/ethnic labels. Some have argued for poor correspondence between these two entities, whereas others have suggested a strong correlation (273).

The two "entities" in question are here "genetic clusters" and "race/ethnicity labels" –or more precisely the groups these labels label. These would be in my terms found genetic populations; and found demographic categories respectively. Still, knowing exactly what the groups are is not necessary for studying them.

> We have shown a nearly perfect correspondence between genetic cluster and SIRE for major ethnic groups living in the United States, with a discrepancy rate of only 0.14% (273).

The authors highlight further that though the distinctiveness of the "major groupings (whites, East Asians, and African Americans)" was perhaps not surprising given previous

research, the distinctiveness of the Hispanic group was not expected due to the genetic distance analysis suggesting a close proximity between the group thus identified and whites.

Objections to the design of the study exist, as well as to how we should interpret these results. It can be argued that these individuals came from uncharacteristically homogeneous SIRE populations within the U.S. For example, the Hispanic group were all residents of a single location in Texas [273], and the authors don't seem to consider whether any limitations in STRUCTURE's model, such as what caused the Kalash to come out in Rosenberg et al (2002) as genetically distinctive, may also have been in operation here.

But what is relevant here is that such a result for the group labeled "Hispanic" would be surprising for a further reason. This would be a group corresponding to an OMB ethnicity and so a group identified explicitly not on the basis of lineage or ancestry, like OMB race would be, but rather on the basis of *either* ancestry *or* culture.

This indicates that a particular embedding of the notion of even ethnicity, which in its ordinary concept may not hold ancestry in its core, could be relevant to genetics. Of course this is a matter of further research, relating to ethnicity specifically. But what is indicated here is that it is not only OMB race but also OMB ethnicity that can be of interest to a genetics context.

This should not be surprising. Founding need not proceed on the basis of an "essential" feature of a concept or a "core" principle. Recall that the founded pen-cap I found as a hair-clip was so founded because of a characteristic only some –though enough– of the pen-caps in my possession had. Similarly, though ancestry need not feature in the core of an ordinary ethnicity notion, or more relevantly of OMB ethnicity the disjunctive "or" in the specification, "origins or culture" indicates that a good proportion of people thus identified, i.e. by OMB ethnicity, may do so on the basis of origins. In any case, it is up to social science research to determine which ethnicities would approximate reproductively closed communities.

3.3.2.2.2 Founding By Proxy

Tang and colleagues do not install SIRE in genetics by calling these common categories 'natural in genetic terms', as Pritchard et al (2000) specify in the specifications of their product, STRUCTURE.

SIRE is not identified with the genetic clusters it is found to label because there is no plausible and conceded upon causal story matching "self-identified race/ethnicity" groups with (just) these genetic clusters that STRUCTURE associates them with. Instead, the four clusters picked out by STRUCTURE, assumed to capture ancient geographical ancestry, "are highly correlated" with SIRE.

For reasons not articulated in this article but articulated in a socio-political and historical context, it is not in the interest of biomedical science to pick up bare-handed these common race categories and install them in the context of genetics. But that doesn't mean there is no founding going on.

Tang et al perform an installation of the categories but they do it by a mechanical arm – by using the proxy 'ancient geographical ancestry' to recruit these ready-made classes as appropriate for answering questions of interest to genetics. And the interests are explicitly interests in confounding and in race:

> This result indicates that studies using genetic clusters instead of racial/ethnic labels are likely to simply reproduce racial/ethnic differences, **which may or may not be genetic.** (…) Therefore, researchers performing studies without racial/ethnic labels should be wary of characterizing difference between genetically defined clusters as genetic **in origin,** since social, cultural, economic, behavioral and other environmental factors may result in extreme **confounding** (274). (Emphasis added)

This is an explicit admission of the fact that these categories are not "genetic" in origin or nature, but to be used, installed, in the context of interests specific to genetics as relevant to answering questions of genetic interest. This is because there is a particular embedding of non-genetic (but still not ordinary) OMB SIRE, which works for the purposes that a genetic category would: it approximates genetic structure –but also possibly 'race' founded in a social context as for example 'socialrace'.

3.3.3 Conclusions

I proposed that founded race concepts such as ‘OMB race/ethnicity’ classifications founded in demography (cf. Appendix 4) are being founded in genomics research. I did so by examining three episodes in what would be a story of finding and founding non-genetic concepts in the context of genetics

1. STRUCTURE is developed from within biostatistics, to found concepts of subjective populations in the context of genetics [Pritchard et al. (2000)].
   This is a founding tool for genetics: it picks up what data labeled under “ordinary populations” is relevant for genetics and articulates it using a genetic model in genetic terms; it is a tool that can be used following a RECIPE to found an ordinary or “subjective” category in the context of genetics.
2. STRUCTURE is used to found populations corresponding to “major geographical regions” or ‘self-identified ancestry’ in the context of genetics [Rosenberg et al. (2002)].
3. STRUCTURE, a found tool for measuring genetic ancestry, is used to found OMB SIRE in genetics [Tang et al. (2005)].

Once more, the possibility of articulating “subjective” populations in genetic terms confirms their appropriateness for studying the relevant ‘questions of interest’. STRUCTURE helps install the respective subjective categories in a spatial context and a context of interests peculiar to population genomics in the case of human populations.

We end up with a certain way of identifying by OMB race/ethnicity classifications, ‘*self-identified* race/ethnicity’, or ‘*SIRE’* which is found to perform relatively well what we thought ancient geographical ancestry did: i.e. measure genetic variation.

And reference to these classes by the jargon of “SIRE” seems of interest. This is a “naming” incident or a “baptism” (which in Greek can mean to ‘dye’ (with a special oil in Christian ceremonies)) that could be a marking out of something of interest and relevance; a new concept of interest and a founded concept for this context. Like a nobility title this too may come with certain privileges but also with responsibilities.

Whether these categories, these found genetic classes turn out to be useful categories for population genetics/genomics remains to be seen. But there is every indication that this question is of interest and not just for geneticists. Common race and ethnicity classes, sequentially founded, are increasingly considered as relevant and appropriate for answering questions of interest to genetics. But some questions interesting to geneticists also interest doctors –and their patients.

Founded 'Race' Can Be Useful

Race concepts in use in these studies are geared to be useful. Distinctions such as 'self-identified'; as opposed to 'genetic' –tagging on 'ancestry' make very clear the causal stories that are supposed to underlie the entity labeled. This is much clearer than talking about 'race'; what Hacking (2006) would call the "biosocial" entity.

Why?

For one it fits current paradigms of causality in epidemiology as distinguishing between environmental and biological causes of disease. Though efforts in the direction of accommodating for complexity are being undertaken in epidemiology such efforts are still in their majority conceived as mixing up two things: environment and genetics.

Second, these distinctions are important for checking processes–such as the checking RECIPE– to work. And we need such processes in order to ensure that a. ordinary concepts and b. scientific concepts stand in some correspondence relationship, in the particular context of enquiry where the correspondence is needed.

Notice that this again will be necessarily local work. Though it might be that a founded concept corresponds well to some ordinary one; the two will not be the same. This is important for understanding what science is and can be doing.

Finally founded race concepts are concepts we can manipulate in science; they are concepts corresponding to entities that may be properly considered as of the type studied in the particular scientific domain. OMB races are not constructs: they are demographic categories.

Similarly self-identified ancestry should it be founded in genetics would not be an ordinary notion; but rather a genetic one; and so would SIRE.

While race and ethnicity are installed in the context of genetics, another attempt to found common race notions happens in epidemiology. What is found in this context is different. Certain differences in health outcomes are found to be stratified according to race/ethnicity classifications. But is this an articulation of 'race' that can found it as an epidemiological object?

3.3.4 Using 'Race' As a Variable In Biomedicine

Race is a prominent category in U.S. epidemiological research. The most glaring health disparities are found to occur between blacks and whites: In 2004 the US death rate for black infants was more than double the death rate for white infants -the difference being almost fourfold for Arizona[121]. Blacks develop heart disease more rapidly and at a younger age than whites[122]. Diabetes mellitus is over fifty percent more common in black adults than in white adults spiking blacks' rates of end-stage renal dysfunction and lower-extremity amputation[123]. Blacks have three times whites' risk of dying from HIV/AIDS [124]; almost half of new HIV/AIDS cases in 2004 were blacks though only about a tenth (12%) of the population identified as black[125].

These are some titles of epidemiological articles:

- Diabetes mellitus, *race*, and socioeconomic status; A population-based study (1996)
- Age-*race* subgroup compared with renin profile as predictors of blood pressure response to antihypertensive therapy (1998)
- *Racial* differences in the outcome of left ventricular dysfunction (1999)

---

[121] The Kaiser Family Foundation, www.statehealthfacts.org. Data Source: Arias E, Anderson RN, Hsiang-Ching K, Murphy SL, Kochanek KD. Deaths: Final Data for 2001. Division of Vital Statistics. National Vital Statistics Report, Vol 52, No. 3, Sept. 18, 2003. Hyattsville, Maryland: National Center for Health Statistics, 2003. Also see Infant Mortality Report in references.

[122] (Root, 2003) 1174

[123] (Brancati et al , 67)

[124] (Root 2003, 1174)

[125] See the Kaiser Family Foundation website, www.statehealthfacts.com:
http://www.statehealthfacts.org/cgi-bin/healthfacts.cgi?action=compare&category=Minority+Health&link_category=HIV%2fAIDS&link_subcategory=New+AIDS+Cases&link_topic=New+AIDS+Cases+All+Ages+by+R%2fE
Data source: Centers for Disease Control and Prevention, Division of HIV/AIDS Prevention-Surveillance and Epidemiology, Special Data Request, November 2005

- Lesser response to angiotensin-converting enzyme inhibitor therapy in *black* as compared to *white* patients with left ventricular dysfunction (2001)
- Survival variability by *race* and ethnicity in childhood acute lymphoblastic leukemia (2003)

These studies examine complex common diseases –diabetes mellitus, blood pressure, heart disease, leukemia – and stratify subjects according to "race" where this is most frequently OMB race/ethnicity.

What is the reason for stratifying results according to race? For one, there is the interest in including race and ethnicity categories in US biomedical research so as to measure and cater to different needs that might exist across racial groups (Epstein 2007). The question is what kinds of "needs" are the ones that can differ across "races" and how do we go about catering for them?

There are two common assumptions about how race and ethnicity categories might matter for measuring and controlling health outcomes. These assumptions are made visible by the way 'race' is featured in the epidemiological discourse, the contexts in which it appears. First, epidemiologists commonly assume that race/ethnicity is associated with social status and social status is thought to causally determine conditions (such as education, access to care, diet, etc. –cf. Appendix 2) that shape health behaviors and health outcomes. OMB Race/ethnicity, assumed to be operating as a proxy for social disparities that causally determine health disparities, can thus be found to track disparate health outcomes. Michael Root is a proponent of this view.

Under such an account, different health needs between racial groups are a trickle down effect of different social needs. Catering to the different health needs of racial groups by this account would involve addressing social needs before they trickle down to medical needs. Or, in epidemiologists' terms, it involves addressing environmental causes of disease.

A second common understanding of how race and ethnicity can track health disparities, involves notions of race and ethnicity as approximating biologically distinctive characters. Though no biologically respectable notion of race is accepted among scientists, race and ethnicity categories are commonly used in the epidemiological discourse as proxies for biologically inherited differences (cf. Appendix 3). Race/ethnicity is assumed to track biological differences between ancient geographical ancestral groups and ancestry is thought to causally determine

genetic inheritance that shapes genetic disease susceptibilities and response to treatment. This is a view on the basis of which Daar and Singer (2005) argue for population-specific pharmacogenetics.

Under this type of account, different health needs between racial groups are in part an effect of inherited genetic differences creeping up, getting triggered or expressed. Catering to the different health needs of racial groups would by this account involve developing race-specific drugs and pharmacogenetic techniques to prevent genetic proclivities from creeping up. Or, in epidemiologists' terms, it involves addressing genetic causes of disease.

Both these understandings of how race matters for health outcomes are called upon to interpret epidemiological findings. And there is a live debate about how to model the causal influence of 'race' on health outcomes[126]. One issue is whether or not to control for environmental risk factors associated with race when studying health outcomes by race/ethnicity categories. Some of the studies whose titles I listed above do not control for social factors such as socio-economic status (SES) or education that might be causally contributing to health outcomes and be differentially distributed across racial groups. In this case differences measured in the health outcomes of racial groups could be due to environmental confounds and so attributing the differences to 'race' rather than say, SES, might be jumping the gun. Some studies try to control for social differences. They select particular measures for social factors that they deem causally relevant for health outcomes and match racial populations with respect to these measures of say SES and education, before looking at whether race effects appear. What happens in many cases is that racial differences in health outcomes persist[127]. And in those cases, the residual effects of 'race' on health outcomes is commonly put down to biologically inherited genetic differences. The thinking is: we took into consideration social differences that shape health outcomes, so what

[126] See for example the exchange between Cooper and Kaufman (2001a) (2001b) and Jones (2001) in the *American Journal of Epidemiology*.
[127] See (1999) study above.

else is left but biologically determined differences and what could be causing such differences but genes –or inherited susceptibilities?[128]

Of course, for this reasoning to be valid, we must have adequately controlled for all the social causes of racial health disparities. Assumptions on at least three levels need to be checked: 1.The initial assumption of what candidate social factors might causally influence outcomes and be racially stratified (Is it SES and education? Is it SES, education and neighborhood? Is it financial distress, education and access to care? Etc.) 2. The selection of tools used to measure these factors (Is SES measured by annual income? Is financial distress measured by the number of times the subject has experienced distress in the last month/year? Etc.) and 3. The techniques used to deploy these tools (Do subjects fill out these questionnaires? Does an examiner fill out the details? Do we use more than one measure of the factor of interest?). All of these assumptions matter!

Baffled by the complexity of 'race' some researchers have proposed to just leave things unpacked. Controlling for environmental factors might be just over-controlling for race.

This dissertation proposes that leaving things unpacked need not be such a bad approach; as long as one is careful not to pick up someone else's bag… That is to say, if race is indeed a complex phenomenon, as it seems to be, why do we need to study it wholesale? This dissertation has argued that there is a clear distinction in discussants' views regarding the two possible foundings of the ordinary notion in biomedical contexts; what I have tagged by the terms 'biorace' and 'sociorace'. What little advice one may give from a philosophical perspective on this issue is, why not examine carefully what these two founded races are or could be?

This does involve but should not rely on only shortcuts to significance, such as "approximating". But developing theoretical, causal accounts about how the founded notions we (researchers) seem to be quite easily operating with map onto the world.

Sure, the "discovered" phenomenon need not look at all like race. But this is precisely why it might be of use in illuminating it. Distance is critical.

---

[128] See Criqui et al (2005) –Criqui et al don't attribute ethnic differences in outcomes of Peripheral Arterial Disease to genetic variability though they ponder its importance. Professor Criqui acknowledged the causal role genetic differences may have in personal conversation (June 2006). Also see Dries, Exner et al (2001)

3.3.5 Found Freedom

The ideal length of a critical distance is hard to figure out; especially hard in the social sciences or in disciplines like epidemiology that straddle the social and natural sciences. Though Root points out a way in which race and ethnicity categories might be significant for biomedical research –as categories which approximate social disparities that influence health outcomes, there is no articulation of precisely which social features that may matter to health outcomes, in what combinations and with what relative weights would shape health profiles and do so along the observed OMB race/ethnicity lines. What makes securing such answers difficult in the case of 'race' has to do in part with disciplinary training.

Geneticists enjoy the luxury of dealing with 'data' on the genetic level –depersonified and often pre-packaged or labeled with only relevant information; clinical epidemiologists do not. Which variables should be featured in the corresponding parametric model of a founding tool for epidemiology? How do we parse out 'individuals' into nicely labeled features, and cluster individuals' variations along these lines in terms "natural" to epidemiology? How do we then cross-reference our results with the common (subjective) OMB race/ethnicity categories available still, and operative, outside the epidemiological context[129]?

These are pressing questions for founding 'race' in clinical medicine and epidemiology.

But answering them is I think in part inhibited by a divergence in attitude that natural and social scientists have towards a problem. What usually happens in social science domains is that common categories are taken up wholesale –or thought to be taken up wholesale. And there is a normative requirement that the social sciences do this in order to study the phenomena at hand[130]. But why should social science bear this brunt? Why are characteristics exhibited on a macroscopic 'human' level demanded to be inviolable by social science theory?

[129] Many thanks to Damien Fennel who raised this worry.
[130] Weber (1949) *The Methodology of the Social Sciences*, Free Press

Here is one answer[131]: faith to social macroscopic phenomena is demanded because these are the phenomena we have to study. As Weber argues, social science may never properly be science, because it has to deal with concepts and phenomena that are just messy; they don't come in the nice measurable forms that natural science is dealt. You can measure length but you cannot measure happiness.

Weber is right: indeed, were social scientists to study race they would have a hard time. But this is not what social scientists do, nor what they should try to do. And I say this because that is not what natural scientists try to do either.

It is a truism to say that social phenomenal are human phenomena and thus harder to control and to study than natural ones. But of course social phenomena are hard to control; but harder than natural phenomena are? Harder than it was to examine the oceans and heavens? I don't think so.

Part of the problem with social science has to do with attitude not with the inherent complexity of social situations. When I studied physics and found out that 'work' is 'distance' times 'force'; I thought how great! How much effort I put in and how far I get is a matter of work. Of course I would never have dared utter that thought in a classroom. I knew it was off bounds. I knew that we were talking about inclined planes and that my thoughts were to be kept to myself. Social science concepts are named in terms that overlap with what Paul Churchland calls "folk" theory too; but it seems to me that there is no such disciplinary shielding in place there (let alone any pride in these concepts and proper naming of them, comparable to how natural scientists espouse their codes –which is perhaps what inspires Churchland to say scientific concepts will eliminate ordinary ones).

Holding social science terms in their proper places and not allowing them to be conflated with ordinary concepts is I think part of the work social science needs to do.

3.3.6 Finding Founded 'Race' as Science in Life

[131] I came up with this answer in conversation with Richard Tutton, so thanks for that.

Arthur Danto writes that one of the concerns of *Fluxus* (the art collective Yoko Ono was a member of) was "overcoming the gap between art and life"[132]. Taking commonplace objects, sounds, gestures, and installing them in the spaces and contexts that make them function as art displaces, along with the objects, our attitude towards them. It comes with the realization that as we change our attitude to given, mundane objects the objects can come to function in different ways.

I argued that a point of "founding" science is overcoming the gap between science and life. I conclude by noting that this would be a two-way process. As life gets founded in science, found science is found as science in life. The capacity to take what is given, and to place it in the context of science, call this physics, psychology or genetics, comes with the responsibility of letting the founded concept function *as science*, in life.

The work of relating this back to an ordinary setting is not dull trivia. Nor is it at all straightforward. It is hard work and in need of competent workers. And it comes with responsibilities. Found science can only be good science if it uses its freedom and is responsible for the "violence" it inflicts. And for this to happen it is necessary that found science is founded with care for each and for the spaces that sustain us.

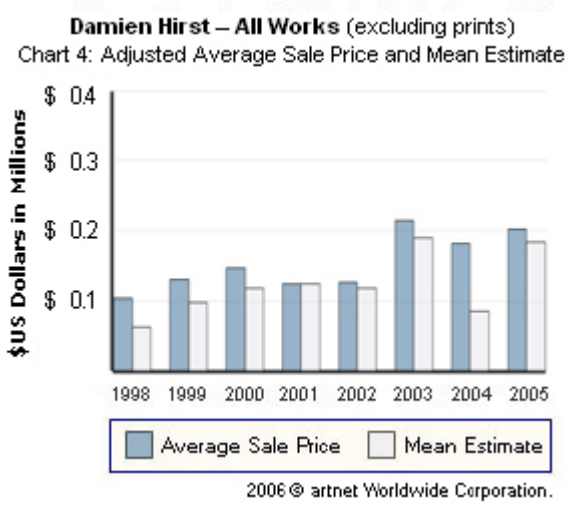


Figure 6: *Placing* what we find interesting in life *in the context of science*

[132] Danto (2000)

Conclusion

This account aimed to accommodate the multi-use of “race” in biomedical research. It did so by underwriting the possibility of such an enterprise by positing the existence of different founded notions of the same ordinary concept in different scientific contexts.

My research indicated that the term is used to refer to multiple concepts of race, which I initially grouped under two second-order race notions: that of race as a social category or ‘sociorace’, and of race as a biological category or ‘biorace’.

My question was: Why and how can this happen in science? How can a concept –even one as historically epistemologically dubious and ethically pernicious as ‘race’ –be used and useful in organizing scientific work? The answer I proposed was that though tagged by the same word, as used in different scientific contexts “race” corresponds to founded race concepts; concepts embedded in different scientific contexts in ways that gear them towards different scientific work in different scientific contexts.

Geneticists and epidemiologists are not to be expected to study races in a scientific context. They cannot. At the same time, they also cannot assume that these common concepts can be reduced to one of these multiple foundings of them. Take founded race notions; these are, to begin with, different kinds of notions: they make use of “race” to get at either a social kind or a biological kind; or even a social or biological statistically significant, meaningful or useful trait. Contrary to our common categories which are commonly discussed as composite entities, carrying forth what Rabinow terms and Hacking (2006) discusses as “biosocial” identities, the categories used in science practice act as either proxies for social groups or proxies for biological differences.

Public health policy targets (social) minority groups to eliminate racial health disparities, epidemiology looks at races as capturing both social differences that matter to health and biologically inherited differences that shape health outcomes and genetics treats races as proxies for biologically inherited variations. Further, what socialrace or biorace notions are articulated,

within each context cannot be *assumed* to match, but rather, have to *be* matched: i.e. re-articulated and founded across contexts.

Revisiting the philosophical debate with the distinction in mind helps articulate the source of the dispute between Root (2003) and Daar and Singer (2005). Daar and Singer (2005):

- Do not demonstrate that what genetic variation occurs between human populations is medically interesting.
- They don't worry about the trans-global applicability of these 'race' categories.
- And, surprisingly for professional ethicists, they don't worry about the socio-political risks for the populations sampled.

Under my account this is because 1. Daar and Singer work with some biorace notion of race to the exception of socialrace and 2. They assume that biorace notions in genetics will match those found to be relevant in epidemiology. This is because they are basing their arguments on research in biological and genetics contexts where founded race notions are found to approximate classes of biological significance.

Root (2003) on the other hand:

- Accepts that race correlates well with disease.
- Rejects the *medical usefulness* of associations between race and genetics because our evolutionary (causal) story does not feature 'race'.
- Proposes self-identified geographical ancestry as a *better proxy* for genetic variation, *but* lacks any data to that end. Tang et al. (2005) actually find 'self-identified race/ethnicity' [surely not a part of our evolutionary story], to be well correlated with ancient geographical ancestry.

This is because Root's working notion of 'race' is one of race as socialrace; it is a founded race notion and one under which races are found social classes. This is because he is to a large extent borrowing his concepts from fields of social science where the categories are reviewed and considered as found status variables.

And consider the problem he mistakenly calls one of using race as a “population variable” and using race as an “individual variable”. Recall that Root accused physicians of mistakenly using race as an individual variable when it was only a population variable. This claim points to where he has seen what he calls “races” used: in epidemiology; and in epidemiology these are found status variables; primarily such as can be aggregated and collected and studied statistically; they are “constructs”. And they are not to be used to pick out real characteristics of these populations; no more than were you to know my socio-economic status would you be able to come look at me and point to it. The fact that race relates to visible physical features is for Root accidental; it only matters insofar as it tracks the targets of racism.

Hacking’s view (2005) was hard to pin down. This is because his understanding of “race” is of race as founded race. Race is first defined as a superficial kind; and so far an ordinary race notion or some other biorace notion would do to describe his concept. But the emphasis on why we classify how we classify is made and race is in conclusion discussed as a socialrace, or a tool for legitimating excessive oppression.

“Race” is multiply used. But not haphazardly so; race concepts in use in biomedical research seem to work as either notions of race as a biological kind or of race as a social kind. This is often expressed as a divergence between authors’ manifest and operative race concepts. But this plurality need not indicate a mistake; it may indeed point to a conceptual modification of heuristic use: what I called a “founded” notion and a case of “found science”.

In a time when political interest in addressing the disparate health needs of US race/ethnicity groups is driving research, the risk of essentializing socialraces and the risk of taking them for bioraces is great. Jumping across scientific contexts without any bridge principles in place, and taking ready-made concepts to be founded when there are merely found is reckless science. To ask and hope to answer the right questions about how ‘race’ relates to ‘health’ we must distinguish between ways race can be founded in science and then to try and translate across founded race notions. We must develop theoretical and pragmatic tools to found concepts within

scientific contexts and to get the founded concepts to be found again and founded across scientific domains.

Part of the problem with race may relate to the vocabulary itself: there is no appropriate distance between epidemiological language and common vernacular to allow the unperturbed founding of these phenomena in science. But part of it has to do with attitude. And though with scientific domains such as physics or genetics folk claims can co-exist with our scientific theories, the discourse of epidemiology is not allowed (and allowing itself) the creative freedoms of genetics. Social science cannot shield its work in a formally closed domain as easily and nurture it to the point of developing its own founding tools. Though there is no good reason why it should not.

I suggest that blind, wholesale "faith to the phenomena" is a mistaken approach founded on an unfounded worry. I do so by indicating that transfigurations of the given commonly take place in scientific practice and have to take place if precision and coherence is to be preserved within particular disciplines but also, for new work to be done. Plus, developing *scientific* theories for life does not eliminate other ways in which we come to theorize life that lie outside a scientific context. Likewise, the unmediated correspondence of biomedical terms and common phenomena expected of biomedicine is unfounded and unnecessary. It is not immediately expected of natural scientific domains that seem to fare quite well even so; so why should epidemiology face these extra constraints?

This account suggests that some intrusion to founded entities "original" place is unavoidable and in fact required to properly found them in a scientific context. If these are human entities or phenomena we study then extra caution should be taken to not violate people's rights. But it is unavoidably work in controlled environments –whether theoretical or physical ones- that must be pursued even here. The extra work of matching the founded objects to (*functions or capacities* of) the common objects must be made visible and taken up as extra work, happening in specific spaces and contexts of interest.

So, is race a genetic class? Or Is it a social marker?

Answer: Race can be a found genetic class and and/or a found social marker.

It depends.

FInding and founding concepts is a process of “articulation” where articulation is separation; a separation that enables the independence and evokes the “life” of what is articulated (Figure 7).

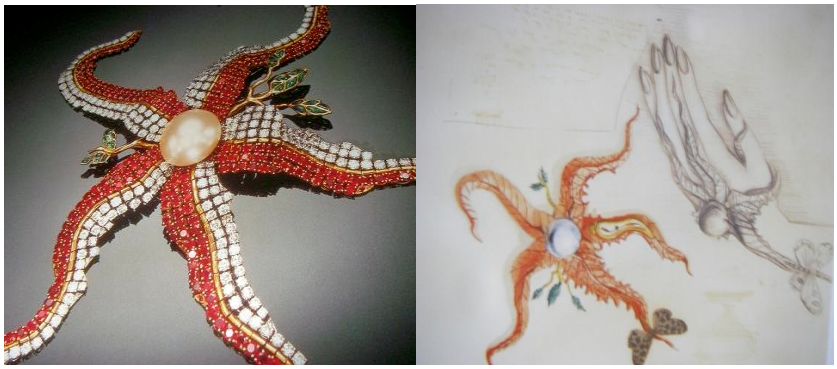


Figure 7: Etoile de Mer (Star Fish) Brooch, Salvador Dali (1950). Pearl, diamonds, rubies, emeralds and gold. Private Collection, Minneapolis, Minnesota, U.S.A.

> A piece that pursues a complex relationship with the body is the Etoile de Mer or Star Fish Brooch created for Rebecca Harkness. This extraordinary work is a technical tour-de-force. Each arm is *fully articulated and can be manipulated into any position, evoking the flexibility of a living creature.* ….
> The Etoile de Mer recalls the fascination with nature and the erotic ambiguity of Art Nouveau, but unlike the hermetic symbolism of the fin-de-siecle, the *piece needs the participation of the viewer in order to create meaning*. In this sense it is a true Surrealist object (Wood 2007, 223). [Emphasis added]

APPENDIX 1

Some Background in Genetics

Race and ethnicity classifications can find a use, in practice, in the context of genetics as markers marking and factors contributing to medically interesting outcomes because 'race' and 'ethnicity' labels can tag bodies and/or bodily matter that is already analyzable in the terms of genetics, some of which is communicated across generations by reproduction, and some of which is capable of causing (or co-causing) disease.

But what is a current symbolic context into which an ordinary concept of race may be founded like?

The point of this section is to give a whiff of what biomedical genetics and genomics might involve. To do this I define some of the terms of what I call "biomedical genetics" and "biomedical genomics" symbolic contexts: contexts in which ordinary race notions I argue are being founded in, and where one can find them as biorace or sociorace notions of race.

I give an outsider's definition of these terms. I am piecing the section together getting information on genetics from publicly available websites like a "talking glossary" made available by the National Human Genome Research Institute (NHGRI) at http://genome.gov/10002096 [Last accessed 01/07/08] and a genetics home reference handbook provided by the U.S. National Library of Medicine http://ghr.nlm.nih.gov/handbook. I want to retain the textbook glossiness of these accounts. And I want to paint the context as much as possible irrespective of 'race'.

I answer these questions:

1. What is DNA and what is the difference between genes and genomes?

2. How are genes related to inheritance and how can one's ancestral lineage be traced using genetic information? [133]

An ordinary race concept manifests a capacity to function in particular biomedical and in particular genomics contexts because some of its features allow it to function in particular ways, some of the times. But what sorts of things interest geneticists and genomicists?

A1.1 Genes and Genomes

An organism's **genome** is the DNA contained in the nuclei of its cells and in its mitochondria, the organelles that generate cell energy (mtDNA)[134].

[133] I am getting information on genetics from publicly available websites like a "talking glossary" made available by the National Human Genome Research Institute (NHGRI) at http://genome.gov/10002096 [Last accessed 01/07/08] and a genetics home reference handbook provided by the U.S. National Library of Medicine that is also very useful for answering further questions. http://ghr.nlm.nih.gov/handbook See also Chapters 1, 2 of Jablonka and Lamb (2005) and pp 45-54 of Rose (2007).

[134] Viruses also have DNA and so a genome but they have no cells; rather their DNA infiltrates the cell of their host. Bacteria are what single-cell organisms are called and they too have a genome.

DNA stands for DeoxyriboNucleic Acid. It is an acid molecule that carries "genetic instructions for making living organisms" [135]. Chemically DNA is a polymer made up of nucleotides that each consist of a molecule of sugar, a phosphate group and a nitrogenous base. The nucleotides are structurally arranged in what looks like a long ladder coiled into a helix, with the sugar and phosphate molecules making the ladder's backbone and the bases, joined weakly in the middle by hydrogen bonds, its rungs[136]. DNA bases pair up in set combinations: Adenine with Thymine and Guanine with Cytosine. As the order of these base pairs affects DNA's function bases, commonly abbreviated to their initial letters **A, T, G and C,** are said to "spell genetic code" or "instructions".

A first step in decoding DNA function is locating **genes**, "the functional and physical unit of heredity passed from parent to offspring"[137]. Genes are interesting because they're DNA containing code for making **proteins** which in turn are very useful for organismic function. Proteins are long chains of 20 different types of smaller molecules called **amino acids** that are indeed quite "protean": they can work as a. antibodies, b. enzymes, c. messenger molecules, d. structural support molecules, e. transport and storage molecules[138]. A gene's code combines four bases to spell 3-letter "words" or **codons** that specify which amino acid is needed at what step in the process of making a protein[139]. Typically a human gene is about 3,000 bases long (the longest is 2,4000,000 bases long) and humans have between 20,000 and 25,000 genes by recent estimates[140].

But not all parts of a gene code for protein; in fact typically less than half of its bases do. The parts that do are called **exons** and those that don't are called **introns,** or junk DNA[141]. And genes (DNA) don't directly code for proteins. Protein synthesis happens in the cytoplasm of a cell via RiboNucleic Acid or **RNA**[142] whose molecule is like a single strand of DNA with Thymine substituted with Uracil. The DNA coded message is "transcribed" onto what is called **messenger RNA** (mRNA) which matures inside the nucleus and exits it acting as a 3-letter "template" for the "translation" of DNA code into the cytoplasm. The two-part process of transcription and translation is called **gene expression**[143] –and the unidirectional flow of information from DNA to proteins is called the "central dogma" of molecular biology[144].

---

[135] http://genome.gov/glossary.cfm?key=deoxyribonucleic%20acid%20(dna) [Last accessed 01/07/08] Steven Hughes (2007) suggests in "Navigating Genomes: The Space in which Genes Happen" in *Tailoring Biothechnologies*, that this common reading of the DNA as "control center" and the language of "gold-rush" mining to describe the process of sequencing obscure the actual macro-dimensionality of "genomic space" the space where genes happen.

[136] http://genome.gov/glossary.cfm?key=double%20helix [Last accessed 01/07/08] The structure of DNA was discovered by Watson and Crick in 1953.

[137] http://genome.gov/glossary.cfm?key=gene [Last accessed 01/07/08] This claim is challenged by studies of inheritance like that of Jablonka and Lamb (2005) that posit the importance of other epigenetic, behavioral and symbolic "dimensions" to inheritance. Also look at Dupre (ms) on different accounts of what genes are.

[138] http://ghr.nlm.nih.gov/handbook/howgeneswork/protein [Last accessed 04/07/08] They a. protect the body by binding to foreign molecules, b. catalyze reactions, c. coordinate processes across different cells or kinds of tissue, d. provide structural support to the body and e. bind and transport molecules to other parts of the body.

[139] http://genome.gov/glossary.cfm?key=genetic%20code%20%28ATGC%29 [Last accessed 01/07/08]

[140] http://www.wiley.com/college/pratt/0471393878/student/animations/dna_sequencing/index.html Humans have much fewer genes than originally expected [Last accessed 03/07/08]

[141] http://www.wiley.com/college/pratt/0471393878/student/animations/dna_sequencing/index.html [Last accessed 03/07/08] Junk DNA is so called because it is of no known function. Viral and bacterial genes have no introns and so they are the shortest types of genes.

[142] http://www.genome.gov/glossary.cfm?key=ribonucleic%20acid%20(rna) [Last accessed 05/07/08]

[143] http://genome.gov/glossary.cfm?key=gene%20expression [Last accessed 05/07/08]

[144] http://ghr.nlm.nih.gov/handbook/howgeneswork/makingprotein [Last accessed 05/07/08], Jablonka and Lamb (2005), p.31.

So to recap DNA is an acid molecule parts of which are called genes parts of which called exons, with the help of RNA, code for proteins. This process is not faultless; mistakes can occur at various points the expression of a gene but corrective mechanisms ensure protein is correctly made.

Perhaps DNA became interesting to look at because of it came in x-marked packets: chromosomes. **Chromosomes** are what a coiled up DNA molecule is scrunched into inside a cell's nucleus: "threadlike 'packages' of **genes** and other DNA"[145]. Different kinds of organisms have different numbers of chromosomes in their cells. Human somatic cells have **23 pairs** of chromosomes, 22 pairs of autosomes and one pair of sex chromosomes, while germ cells (eggs or sperm) contain half that number, so 23 chromosomes for humans. Each parent contributes one chromosome to each pair so humans get half their chromosomes from their mothers and half from their fathers. Though a mother's egg contributes mitochondria and so mtDNA to a maturing embryo[146], which in the case of humans contains 37 genes[147].

All of introns, exons, genes, genomes and chromosomes are DNA individuated by their function and/or spatial structure. Genome **sequencing** is figuring out the order of base pairs in the DNA we call the genome (so chromosomes and mtDNA).

A1.2 Sequencing and Mapping

The Human Genome Project (HGP) started in 1990 and aimed to give a complete sequence of the human genome (including genomes of other organisms)[148]. It was an endeavor involving 18 countries that was heavily funded by the U.S. government and the U.K. Sanger Center. Sequencing the human genome was "completed" in April 2003 after 99% of the protein coding parts of the human genome was sequenced to 99.99% accuracy[149]. The human genome has about three billion bases and is by recent estimates 99% the same in all of us[150] and surprisingly similar to that of simpler organisms.

The HGP would still be going had DNA sequencing techniques not rapidly evolved[151]. Up until the late 1980s the most popular technique for sequencing DNA was protein sequencing that was very slow and expensive. It could take a year or more to sequence a protein of about 500 amino acids and sequencing a gene could take 10 years! Sequencing is now automated. Sequencing a protein takes a few days and the cost per genotype went down from 50 cents in 2002 to a tenth of a cent in 2008[152].

How does sequencing work? *Dideoxynucleotide sequencing* (commonly called dideoxy or Sanger sequencing) is a common DNA sequencing technique. The technique was invented in 1977 by Frederic Sanger and his colleagues Paul Berg and Walter Gilbert for which they were awarded the 1980 Nobel prize in chemistry –making this Sanger's second Nobel prize in chemistry. The technique is based on using nucleic acids as opposed to proteins in the process of

---

[145] http://genome.gov/glossary.cfm?key=chromosome [Last accessed 01/07/08]
[146] http://ghr.nlm.nih.gov/handbook/inheritance/inheritancepatterns [Last accessed 06/07/08]
[147] http://ghr.nlm.nih.gov/chromosome=MT [Last accessed 05/07/08]
[148] http://www.ornl.gov/sci/techresources/Human_Genome/hg5yp/ [Last accessed 08/07/08]
[149] ibid. [Last accessed 08/07/08]
[150] http://ghr.nlm.nih.gov/handbook/basics/dna [Last accessed 03/07/08]
[151] See http://www.ornl.gov/sci/techresources/Human_Genome/hg5yp/ [Last accessed 08/07/08] for HGP progress reports. See also Kitcher's (1996) pessimistic predictions regarding the achievement of the Human Genome Project's goals.
[152] These are figures presented by Francis Collins to members of House of Lords during a two-day educational seminar in June 2008. See "NIH Scientists Introduce British Lords to Genomic Medicine" http://www.genome.gov/27526872 and materials available there. [Last accessed 08/07/08]

sequencing[153] and the fact that it can be automated played a key role in speeding up (and lessening the cost of) DNA sequencing in the later half of the 20th century. I briefly describe this procedure because I think it is fascinating, because it goes on all the time, and because I use it as an example to make a philosophical point in the thesis (Chapter 2 Section 2.3.1.1).

The technique works as follows[154]. First, a cell's nucleus and membranes are broken open either mechanically or chemically so that the DNA can be extracted. The DNA thus obtained is still covered with cell debris so it is first purified by precipitating it and cleaning out DNA-binding proteins. Large pieces of DNA such as chromosomes are then broken up into smaller pieces and inserted, via appropriate vectors, in bacteria or viruses that are then cultured and multiply to give enough material to later parallel sequence the same DNA fragment. Each line of multiplied DNA is called a clone. *Dideoxy sequencing* is then taken up for each member of each cloned line, which means just this.

Clones of each DNA fragment go through a reaction with four main steps: 1. strand separation, 2. primer annealing, 3. extension, 4. termination. First, DNA is extracted from the bacteria and heated, so its base pairs break up and we're left with two strands of it: one of which is called (because it acts as) the "template" strand and the other one the "complementary" one. The two separated strands are then mixed in with oligonucleotides called "primers". These are small single strands of DNA that are hoped to be complementary to and so able to latch onto, or "anneal", the beginning of the template strand. In theory the longer complementary DNA strand can just stick back onto the template, so the mixture is cooled to help the shorter primers reach the template strand before the complementary strand does. In the third step, using the primer as a guide, DNA polymerase (a protein acting as a catalyst for DNA replication) fills in missing bases in the rest of the template strand. Thus the primer is extended via the polymerase. The process stops randomly because end-tags are introduced in the mixture in the right proportions. These are dyed dideoxy nucleotides, one base long and color-coded by base that can latch onto the replicated chain but are chemically modified so that no more nucleotides can bind to them.

After copying terminates, filled-in DNA chains are then again broken up, the negatively charged fillers kept and forced to pass through a tiny filament at the end of which is a positive charge in a process known as "capillary electrophoresis". The filament is full of viscous fluid so shorter pieces that meet less resistance in the fluid sink through the liquid first, so in effect this process sorts DNA strands by length: the strands go down the capillary one by one each one base shorter than the following one[155]. A laser beam at the end of the filament is then shone on the emerging strands and excites the dye in their end tags, which fluoresces and is thus detected by a photocell –which is hooked up to a computer that records the frequencies of light emitted by the end tags in an electroferogram. The order of colors thus detected corresponds to the order of DNA bases in the (cloned) fragment of the DNA sample, which can thus be proclaimed sequenced.

---

[153] Craig Venter, a UCSD undergraduate and PhD, and founder of *Celera Genomics*, the private company that published results of the first complete human genome sequence along with the HGP developed a technique called Expressed Sequence Tags (ESTs) which sped up genome coding further by allowing the sequencing of protein coding regions of DNA.

[154] http://www.wiley.com/college/pratt/0471393878/student/animations/dna_sequencing/index.html [Last accessed 01/07/08]. This is taken from an exercise on DNA sequencing accompanying an introductory biochemistry textbook for college courses titled *Essential Biochemistry*, by Charlotte Pratt and Kathleen Cornely, Wiley and Sons (2004).

[155] I suppose that fluid resistance varies more significantly by length than the attractive force these strands experience: i.e. if longer strands are more negative than small ones and so experience a stronger attractive force, the force must still not be strong enough to speed them up through the fluid and make them come out before short strands. I am also not sure why it is said that the process will randomly stop at a possible base point just once (and so strands will each differ by one base length)– perhaps the number of parallel replications is large enough for a statistical law to apply here, or perhaps this is a simplification for the purposes of teaching this technique. Comparing sequences at the next stage can hopefully pick up some mistakes.

At this point biostatistics techniques become important. Mathematical tools are used to re-assemble codes obtained by multiple Sanger sequencing of multiply fragmented and multiply copied DNA. Re-assembly used to proceed by pair-wise comparison tests but it too has now sped up using faster clustering algorithms[156]. Information thus obtained on genome sequences and assemblies is stored in bioinformatics databases, cross referenced and managed by various institutions[157].

Sanger sequencing is a typical sequencing technique. But sequencing DNA is not the only way to obtain information from DNA molecules. Genetic information is communicated by the positions of entities of interest other than bases in genetic spaces other than a DNA chain. Getting this information is called DNA "mapping".

"Maps" are schematic representations of DNA nucleotide structures that come in three types:

- A "gene" or "linkage map" specifies genes' relative positions on a chromosome and the relative distance between them[158].
- A "physical map" specifies the physical locations of genes or markers on chromosomes. Physical maps are useful for DNA sequencing and when searching for genes responsible for diseases using "positional cloning" techniques[159].
- Finally a "cytogenic map" is what a chromosome looks like when stained and examined under a microscope. A cytogenic map has regions called light and dark bands that give each chromosome the distinct appearance which a clinical test called a karyotype uses to look for chromosomal alterations[160].

A variant form of a gene on a chromosomal location is called an **allele**[161]. Different alleles of the same gene are responsible for the observed variation in inherited Mendelian traits like hair color or blood type. If an individual inherits different alleles for a gene from its parents and one is more often expressed than the other the alleles are called "dominant" and "recessive" respectively.

Until recently it was thought that both chromosomes in pairs of autosomes contributed in a similar fashion to a person's genome and genome sequencing was haploid containing only 23 of the chromosomes. The first diploid genome to be sequenced in spring 2007 shows that this is not the case[162]. The authors claim, "These data depict a definitive molecular portrait of a diploid human genome that provides a starting point for future genome comparisons and enables an era of individualized genomic information"[163].

A1.3 Markers and Factors

---

[156] Venter describes this in a speech he gave at UCSD in June 2007 upon being awarded the 2007 Nierenberg Award. See or listen to Venter's acceptance speech: http://www.uctv.tv/search-details.asp?showID=13024 for video and http://podcast.uctv.tv/mp3/13024.mp3 for audio [Last accessed 01/07/08].

[157] http://genome.ucsc.edu/index.html?org=Human&db=hg18&hgsid=109174337 [Last accessed 05/07/08] An example of such a database.

[158] http://www.genome.gov/glossary.cfm?key=gene%20mapping [Last accessed 02/07/08]

[159] When geneticists speak of clones they mean copies of particular pieces of DNA (usually genes) –not exact genetic copies of organisms.

[160] http://www.genome.gov/glossary.cfm?key=cytogenetic%20map [Last accessed 02/07/08]

[161] http://genome.gov/glossary.cfm?key=allele [Last accessed 02/07/08]

[162] The genome was Craig Venter's and the article is openly available on PLoS Biology: http://biology.plosjournals.org/perlserv/?request=get-document&doi=10.1371/journal.pbio.0050254&ct=1 [Last accessed 02/07/08]

[163] http://ucsdnews.ucsd.edu/newsrel/health/11_06_DNA.htm [Last accessed 01/07/08]

Sequencing and mapping DNA should only be of interest if genomic space can change. And it does -frequently and sometimes with consequences to organismic health. Permanent structural changes in DNA -which amount to an exchange, deletion or insertion of bases in the DNA sequence- are called **mutations**[164]. Mutations occur very frequently in our DNA -on average one mutation every time a cell multiplies[165] and cells have special systems for repairing such mistakes even if they are usually of no consequence to an organism. Mutations that are more than 1% frequent in a population are called **polymorphisms**.

There are two uses a mutation (or anything for that matter?) can have:

- it can be a symbolic "marker" –when used as input for a tool or operator
- or a causal "factor" –when used as a tool or operator

A mutation can function as a marker of a feature of interest within a context of scientific inquiry and function as a causal factor within a biological system. Let us review the two types of function that a mutation can have in biomedical contexts of inquiry[166]

A **genetic marker** is a "segment of DNA with an identifiable physical location on a chromosome whose inheritance can be followed"[167]. Markers can be genes but also DNA sections with **no** known function. For example, DNA polymorphisms called Single Nucleotide Polymorphisms (shortened to **SNPs** and pronounced "snips") are common, tiny variations in the DNA sequence that amount to a misspelled base in the sequence that are often used as markers. They occur at a frequency of 1 every 1,000 bases in human DNA[168] and are usually of no importance functionally but still of genetic interest. This is because bits of DNA that are close together on a chromosome tend to be inherited together and so by following the inheritance patterns of markers one may be following inheritance patterns of genes that have not yet been identified but whose approximate locations are known[169]. For example say that a gene if misspelled could increase your risk of high blood pressure. But researchers don't know where in the chromosome that gene is. They can compare SNPs of people who suffer from high-blood pressure and those who don't and if a particular SNP is more common among people with the disease the location of that SNP can act as a pointer to the location of a gene involved in the disease[170]. Chromosomal regions with genetic variants whose inheritance is linked are called **haplotypes**[171]. Mapping haplotypes instead of the locations of particular genetic mutations is seen as shortcut to studying human genetic diversity as many regions of humans' chromosomes have only a handful of haplotypes.

---

[164] Mutations are of five types: 1. deletion, 2. duplication, 3. inversion 4. insertion and 5. Translocation.

[165] http://www.insidecancer.org/ [Last accessed 02/07/08] Bert Vogelstein, M.D. a Howard Hughes Medical Institute Investigator and the Clayton Professor of Oncology and Pathology at Johns Hopkins University gives an interesting description of how cancerous cells develop. His research focuses on the identification and characterization of genes that cause colon cancer. This has led to the discovery of the APC gene – the "gatekeeper" in colon cancer development.

[166] Nikolas Rose picks out a paradox in his paper (ms) regarding the use of 'race' which can be explained as an example of this dual function of a thing as thing (factor) and symbol (marker). He points out that in many biomedical contexts "races" can be deemed fluid and dynamic entities and yet how these categories can be used when performing ancestry tracing tests as useful and meaningful, fixed labels; tags for more complicated information about individuals. I think this is an effect of a shifting in attention between different contexts of enquiry which demands holding fixed certain variables in order to express another variable in terms of those. Rose, N. (ms) "Race, risk and medicine in the age of 'your own personal genome'"

[167] http://genome.gov/glossary.cfm?key=marker [Last accessed 02/07/08]

[168] http://www.genome.gov/glossary.cfm?key=single%20nucleotide%20polymorphisms%20%28SNPs%29 [Last accessed 04/07/08]

[169] http://genome.gov/glossary.cfm?key=marker [Last accessed 02/07/08] This is the technique used by the HapMap project which studied for regions of the DNA rather than particular sequences.

[170] This example of course starts with the assumption that a gene is involved in the disease.

[171] 'Haplo' means 'simple' in greek.

The **International Haplotype Map Project** (HapMap for short) is identifying haplotypes in "populations with African, Asian and European ancestry" [172]. HapMap is "identifying the basis for a large fraction of the genetic diversity in the human species"[173]. Haplotypes are being identified along with "tag" SNPs that identify each haplotype uniquely. The tag SNPs can thus be genotyped and used in reverse as markers of haplotypes in a person's DNA. The number of tag SNPs in search of is estimated to be about 300,000 to 600,000 which seems like a lot but is much fewer than the some 10 million common SNPs in the human genome.

It is envisioned that tag SNPs will be useful for locating genes for medically important traits. Instead of identifying all SNPs associated with a disease like high blood pressure in all subjects in a study a researcher should only need to genotype tag SNPs in each individual to figure out their haplotypes. And knowing individuals' haplotypes enables researchers to either focus on candidate genes associated with the disease or look across the genome to find regions associated with it: "If people with high blood pressure tend to share a particular haplotype, variants contributing to the disease might be somewhere within or near that haplotype"[174].

Though the terms "race" and "ethnicity" only come up under "ELSI" (short for ethical, legal and social issues) on the HapMap project's website note that that race concepts are involved in the project's design. Haplotypes are sought using populations that could be roughly identified as racial. What would fall out of this method is that disease associated tag SNPs and haplotypes will also come with a racial label on, should they choose to use it.

A mutation in genetic material that is inherited across a particular lineage is also used to individuate a lineage (mark it) and subsequently trace it to a geographic place of "**origin**" [175]. A technique developed by Luca Cavalli-Sforza in the 1970s and improved since involves sampling the DNA of populations from different parts of the globe and comparing patterns in their DNA markers.

Ancestry-tracing methods have various shortcomings, which are relevant to the discussion of race in genetics. For example the "origins" of these lineages are in practice places where populations who have particular genetic markers resided when their DNA was collected. Supplemental knowledge saying these have been reproductively isolated, speak the same language, are indigenous etc. is needed to justify the claim that this was the place of "origin" for particular lineages. Further, genetic markers are a very small proportion of inherited genetic material and not portions that are straightforwardly biologically meaningful. Finally most ancestry-tracing services use markers on mtDNA or Y chromosomes which are shared along one's maternal and paternal lineages respectively and as the mothers of your mother and the fathers of your father are only very, very, very few of your immediate blood relatives this type of analysis captures only a small proportion of one's "ancestry"[176].

---

[172] Information is at: http://www.hapmap.org/whatishapmap.html

[173] Ibid.

[174] Ibid.

[175] Geneticist Spencer Wells nicely illustrates this process in the movie "The Journey of Man: A Genetic Odyssey" [PBS]. http://www.youtube.com/watch?v=ybji0axp6s0&feature=related [Part 2 of 13, Last accessed 28/06/08]. The movie is quite interesting in itself: it follows this fit blonde man, geneticist Spencer Wells, who bids his family goodbye in the first few scenes and sets out to follow in the steps of his ancestors by traveling to distant locations across the world where various human populations live and trying to piece together the story of how ancient human migrations have created the present distribution of genetic material across the globe. The narrative is ideological in a pejorative sense as a disproportionate emphasis is placed on the epistemic cache of the genetic ancestry-tracing techniques used by Wells as opposed to linguistic features, fossil record readings, verbal accounts etc that are all collected as relevant evidence for the final verdict. It paints a picture of a related family of Man on the basis of collected bodily material and DNA while non-genetic features of the populations discussed are limited to similarity in looks –and how receptive they all are to their long lost cousin. Still, even if it's not the whole story, it is a story worth hearing.

[176] Troy Duster's talk at the 2006 MIT Center for Diversity in Science and Technology conference emphasized this problem.

What genetic structures serve as markers of interest in the context of a theoretical investigation may also be factors of causal import to an organism's **function**. Given the importance of DNA for cell function and protein production and the significance of those for the life of the organism it is easy to imagine genetic and genomic structure affecting the function of an organism in very radical ways. Genes do not cause disease but mutated genes may and such mutations could be inherited or acquired in the organism's lifetime[177]. Mutations may be inherited in different patterns depending on where in the genome the mutation is (X-linked, autosomal or mitochondrial) and whether it is dominant or recessive[178]. Genetic counseling aims to help patients estimate their risks of genetic disorders and gene therapy that involves replacing, manipulating or supplementing non-functional genes[179] with healthy ones is tested in humans with incurable conditions[180].

The term 'gene' was coined by Danish botanist Wilhelm Johannsen while trying to formulate a biological notion of heredity[181]. Already in 1911 Johannsen proclaimed "heredity may then be defined as the presence of identical genes in ancestors and descendants…"[182] and still a gene is defined by the NIH as "the basic physical and functional unit of heredity"[183]. Johannsen was working with pure lines of plants (bred by self-fertilization from one individual) and noticed that plants of the same line would not vary much. Also when the extremes of variation were bred separately –say the tallest ones or shortest plants- trends would not persist but average back to the pure line mean. This observation led Johannsen to define the concepts of **genotype** and **phenotype** which are still widely used to understand disease etiology.

The terms 'geno-type' and 'pheno-type' are compounds of the greek 'genomai' to be born/ become or 'phenomai' to appear. Genotype is defined as the "genetic identity" of an individual; it need not manifest as outward features[184]. It rather captures the inherited potential of an organism that can be realized differently depending on how the organism is raised. Phenotype is defined as made up of observable characters like weight, eye color or blood pressure and need not be fully genetically determined. Extreme phenotypic variations in the same pure line are explained by non-heritable environmental causes and the non-reproducibility of these differences once the extremes are further bred is explained by their shared genotype [185].

And the completion of the HGP burst any bubble of hope geneticists had for decoding life via genetic information. To quote Craig Venter:

> "We found that we could not define the life of even the simplest cell based on its genetic code alone. Without understanding the environment in parallel you can't define life. (…) All of us, every life form, is defined by the environment as much as the genetic code. We

[177] http://ghr.nlm.nih.gov/handbook/mutationsanddisorders/mutationscausedisease [Last accessed 09/07/08]
[178] http://ghr.nlm.nih.gov/handbook/inheritance/inheritancepatterns [Last accessed 06/07/08]
[179] http://genome.gov/glossary.cfm?key=gene%20therapy [Last accessed 06/07/08]
[180] http://ghr.nlm.nih.gov/handbook/therapy/genetherapy [Last accessed 09/07/08]
[181] Jablonka and Lamb (2005), p.28. See also Mueller-Wille and Rheinberger eds. (2007) *Heredity Produced* for histories of the notion of heredity.
[182] See Jablonka and Lamb (2005), p 28.
[183] http://ghr.nlm.nih.gov/handbook/basics/gene
[184] http://genome.gov/glossary.cfm?key=genotype [Last accessed 05/07/08]
[185] Is the genotype genetically determined? In particular, does the central dogma of molecular biology stand or does it like most dogmas face challenges? Could information flow from protein to RNA and to DNA? Are there other mechanisms by which information can affect the genotype? It depends on how information flow is conceived. There are trends within biomedicing such as proteomics (defined as functional genomics) that shift their attention to other levels of cell organization. And philosophers of science also have interesting responses. See Jablonka and Lamb (2005) *Evolution in Four Dimensions –Genetic, Epigenetic, Behavioral and Symbolic Variation in the History of Life* MIT.

> can measure the genetic code. Measuring the environment is a much more complicated task"[186].

The NIH Reform Act of 2006 (Public Law 109-482) mandated the better conceptualization of scientific questions so "The translation from discovery to patient care will be better facilitated"[187]. NIH reforms announced two years after the passing of the act as falling within its mandates include trans-NIH research funded by a Common Fund which the Act established for research cutting across boundaries of different NIH centers and organizations.

Projects funded under trans-NIH research and said to be building onto and supplementing the knowledge provided by the Human Genome Project include:
the *Human Microbiome Project* (HMP) which aims to study microbial cells which inhabit the bodies of healthy humans, in an estimated proportion of 10 microbial cells to one human cell and are largely unstudied: "These hidden communities of cells are the unexplored planets of human biology"[188]. The *Epigenome Project* is to determine influences on gene function, such as environmental factors which "regulate or turn genes on and off". This is a project consisting of a series of integrated initiatives launched by the NIH in September 2008. Some of the expected outcomes of the project are a map of the epigenomes of normal human cells against which to reference those of diseased cells, a Data Coordinating Center expected to enhance data sharing world-wide and the discovery of what regulates epigenomic structure.

---

[186] See or listen to Venter's acceptance speech for the Nierenberg prize: http://www.uctv.tv/search-details.asp?showID=13024 for video and http://podcast.uctv.tv/mp3/13024.mp3 for audio [Last accessed 01/07/08].

[187] Circulated by the U.S. Department of Health and Human Services, National Institutes of Health (NIH) News, NIH Office of the Director (OD) http://www.nih.gov/icd/od/ Tuesday, September 9, 2008 under "THE NIH REFORM ACT OF 2006: PROGRESS, CHALLENGES, AND NEXT STEPS" Statement of Elias A. Zerhouni, M.D., Director, National Institutes of Health, U.S. Department of Health and Human Services

[188] Ibid. The analogy is amusing given that one of the people nominated to serve on the scientific management review board for this project is former Nasa administrator Dan Goldin.

APPENDIX 2

Nature Genetics Supplement: Volume 36, no 11, November 2004

I argue that medical and philosophical accounts I have examined conceive of race in two ways: a. as a social kind and b. as a biological kind. I call concepts that correspond to these two types of race notion 'sociorace' and 'biorace' notions respectively. This is problematic as notions are conflated both in medical practice and in philosophy. However, the form of philosophical arguments I have examined is different from the form of arguments in medical discussions. Biomedical articles acknowledge problems with the ontological status of races but they mostly push through and argue for the use of these classifications in practice. Philosophers put up STOP [or GO] signs about using "race" in medicine based on what ontological status they think the categories have. [And none of them want to re-label these categories as I do.] This discrepancy and its foundations are explored in SYMPTOMPS.

The articles reviewed here are taken from a major and mainstream U.S. biomedical journal. And the analysis offered is targeted to expose the discourse but also support my more general argument.

On how authors use 'race' I argue that

> there is a multiplicity of understandings of what "race" is, a fact acknowledged by medical researchers and practitioners.
> These multiple understandings are already seen as organized along the axes of social/biological and environment/genetics. I use distinctions between different race concepts to organize different race concepts.

On how authors answer the normative question of whether/how race should be used in medical genetics I argue that

> it is overall conceded that ordinary races ARE NOT genetic populations but rather socially constructed classes.
> It is also argued that the term 'race' SHOULD be used in the medical genetics literature. This is 1. to avoid confounding non-genetic and genetic causes responsible for variation measured across the groups of people we call 'races' and 2. to explore the genetic basis for health disparities commonly measured across these groups.

Race specific medical practice has, does, and so can, proceed despite this multiplicity of understandings of "race". The ontological grounds for this practice are explained in Chapter 3, Diagnosis, in terms of found science concepts.

This is when the discussion on race and genetics became loud and clear[189]. On May 15, 2003 a meeting on the topic of "Human Genome Variation and 'Race' –The State of the Science" was held at the National Human Genome Center (NHGC) at Howard University[190]. The timing and location of the meeting are telling. The Human Genome Project (HGP), significantly subsidized by the U.S. federal government, started in1990 and was completed only the month before the

[189] See Rose (2007) Chapter Six. Epstein (2007) especially Chapter Ten offers a comprehensive account which clarifies that debates about using race categories in biomedicine in the U.S. began in the early 1990s, which Epstein says is quite late, given the murky nature of these categories had been already touted in the 1970s after Richard Lewontin's (1972) study of blood group polymorphisms.
[190] http://www.genomecenter.howard.edu/index.htm

Howard meeting. And Howard is a historically "black" university suggesting the meeting should be "endorsed" by a race visibly abused by racism in the United States[191]. I review a set of the meeting's proceedings published as a *Nature Genetics* Supplement on "Genetics and the Human Race" in November 2004. Information included (and omitted) represents the point of view of scientists working in a major genetics research initiative.

If anything, the staging of this public discussion and the coverage it received validated the salience of "race" in public discussions of scientists about science. The scientific community shared the position captured in the statement on race issued by the American Anthropological Association five years earlier in 1998:

> The racial worldview was invented to assign some groups to perpetual low status, while others were permitted access to privilege, power and wealth. The tragedy in the United States has been that the policies and practices stemming from this worldview succeeded all too well in constructing unequal populations among Europeans, Native Americans, and peoples of African descent. Given what we know about the capacity of normal humans to achieve and function within any culture, we conclude that present-day inequalities between so-called racial groups are not consequences of their biological inheritance but products of historical and contemporary social, economic, educational and political circumstances[192].

The explicit attempt to relate race to variations recorded in the just sequenced human genome was a move visibly opposing the eviction of "race-talk" from respectable scientific forums.

A2.1 Royal CDM, Dunston GM. "Changing the paradigm from 'race' to human genome variation"

The founding director of the NHGC at Howard, Georgia Dunston, and then Director of the GenEthics Unit at Howard, Charmaine Royal, open the discussion. They claim that knowledge of human genetic variation is "forcing a paradigm shift in thinking about the construct of 'race'" (S5). It's not clear what the new paradigm is; though it's clear *genetics* is involved in the paradigm shift.

Royal and Duston report differences documented in the "structure of sequence variation" of African and non-Africans' genomes. What these differences mean isn't specified, though authors say knowledge derived from the HGP challenges the validity of standard race/ethnicity categories as fitting for biomedical and genetics research (S5). But this attitude changes. The authors say research on health disparities between ethnic/racial groups attributes them "almost exclusively" to social, cultural and economic causal factors rather than genetic ones and they suspect this is because of 1. a "general perception" that genetics plays a minor role and 2. technological limitations in studying the genetics of complex common diseases prior to the HGP. They imply things are different now.

It is the position of the National Human Genome Center (NHGC) that genetic knowledge *can be used* for better or worse: it could exacerbate health disparities if used only by a powerful sect of society to combat rare diseases[193] or if used to stigmatize groups that diverge from a single physical ideal. But the same knowledge can effectively be used to "eliminate" health disparities, the authors claim. How? By:

---

[191] http://www.ornl.gov/sci/techresources/Human_Genome/home.shtml [Last accessed June 23, 2008]. Various completion dates for the HGP are given there.

[192] http://www.aaanet.org/stmts/racepp.htm [Last accessed December 10, 2008]

[193] It is an uncharitable reading I am sure but Tay-Sachs came to mind here along with anti-semitic stereotypes.

- targeting common diseases in the *least healthy* groups in society,
- revealing the *causes* of health disparities,
- *sharing* the benefits of this knowledge with vulnerable groups and
- understanding the message of this effort as valuing variation as "an instrument of *self-discovery*" (S6).

Points 3 and 4 above refer to the use knowledge in general can have to empower and educate[194]. 1 and 2 are more specific. They posit genetics knowledge can help figure out why the common disease burden is unequally distributed over different social groups. And it is more specifically added the NHGC at Howard University aims to understand the "causality, prevention and treatment of diseases common in African Americans and other African Diaspora populations" and do so through planned research in Translational Genomics in the African Diaspora studying relationships between "gene-environment interactions, complex traits and health disparities between racial or ethnic groups" (S6).

By this point the authors clearly suggest the construct, race, may be genetically useful. Racial and ethnic group categories are to be investigated as categories that could track genetic causes of disease and which can help distribute biological, genetics knowledge equitably across interested parties. Could this work?? Well, I don't know. But these geneticists seem to think it could.

The question here is not whether they're right. But how they can be thinking what they are about 'race'. Manifestly Royal and Duston use a sociorace race concept: they define races as constructs. A sociorace race concept would explain why standard race/ethnicity categories would be inappropriate for genetics research. But their operative race notion is one that imputes to race a biological reality –even if this is in what Hardimon calls the basic sense (-Let us then study 'dragons'). They are saying races are labels that pick out something biological –even if not "essential" or "naturally evolved" or such. They are using a biorace race notion.

To see the difference, reason by contradiction. Say race was thought of as a construct. Now compare this with a label we may more confidently call a construct, like 'marital status' or 'ZIP code'[195]. These could also –like race- happen to tag onto living humans, which happen to have genes or such biomedically interesting features. But we don't see research centers studying the genetic basis for 'ZIP code'-associated health disparities.

Should 'race' categories be used to approximate medically interesting human genetic variation?

There is no clear answer. Races are referred to explicitly as constructs but also as constructs fitting to discuss in genetics. In responding on behalf of the NHGC to allegations that it perpetrates "race-based science" authors claim the term 'race' as applied to humans, *is incorrectly used*: "Traditional 'racial' designations in humans are not bounded, discrete categories but are fluid, socially defined constructs that have some poorly understood correlations with various biological elements and health outcomes" (S6). The authors though hope their work will illuminate the interaction of genes and environment as well as "cultural and other psychosocial factors that contribute to common complex diseases"(S6).

[194] Though one must wonder whether reference to "self-discovery" tools is pitch for a booming ancestry-tracing industry. A link on the NHGC's webpage for example leads to http://africanancestry.com/ [Last accessed …]. See Rose (ms) for further discussion on ancestry-tracing sites and the tools they use.

[195] These are U.S. postal area codes which often associate with more or less affluence. The series "Beverly Hills 90210" branded itself by such a ZIP code.

A2.2 Collins F, "What we do and don't know about 'race', 'ethnicity', genetics and health at the dawn of the genome era"

Then Director of the National Human Genome Research Institute (NHGRI), Francis Collins acknowledges 'race' and 'ethnicity' are ill-defined terms that carry complex socio-cultural connotations. Still, he claims, genetic variation can be used to make a pretty accurate estimate of the geographic origins of an individual whose grandparents all came from the same part of the world. And as ancestral origins are correlated with self-identified race or ethnicity there could be some biological basis for race, he reasons. However, the vague boundaries between populations, the fact that many people have ancestors from different geographical locations as well as the non-genetic connotations of race imply that the biological basis for race is quite blurry.

The second question he asks is **whether health disparities between different races have any genetic basis**. Though there will be many cases where differences in culture, diet, socioeconomic status, access to care and other environmental factors are responsible for health disparities, there are certain recessive disorders like Tay-Sachs or sickle-cell disease associated with specific alleles that are distributed unevenly among human populations. Collins gives a diagram on "Interconnections between self-identified race or ethnicity and health status" to clarify his view (Figure A1).

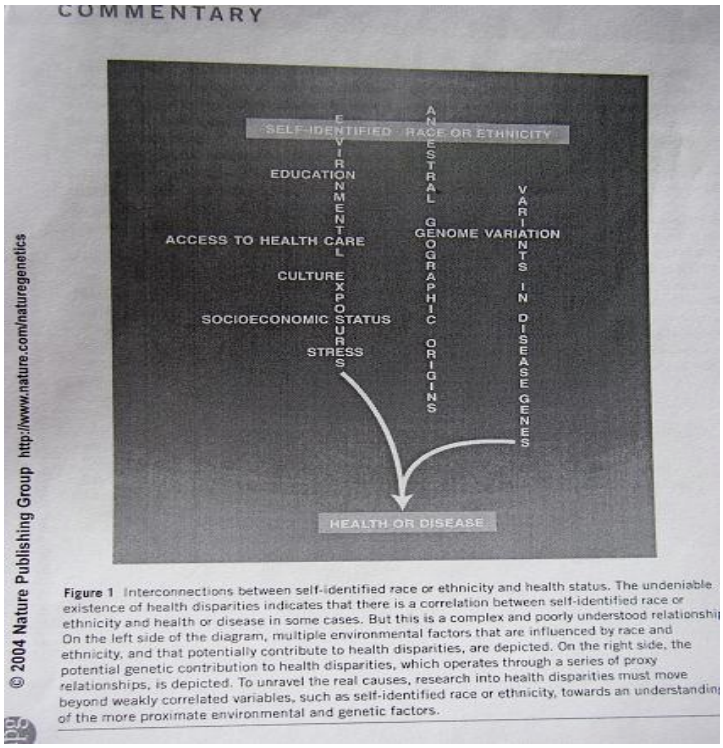


Figure A1

Figure A1 is not a causal graph and there are no arrows on it nor a sense of directionality but for the end bit. Rather words are linked up like words of a crossword puzzle via letters they share[196]. 'Race' is linked to 'health or disease' but the relations this scheme represents are weak

[196] I like Collin's heuristic of linking up words. I think it destabilizes the notion of a fixed, locatable cause represented in terms of a node in a causal graph. What would be "nodes" here are words and sticking other words through them is unsettling enough when you try to read them to communicate visually the complexity of studying causal relations between these categories. The form of this diagram also reminds me of what epidemiologists call "multicausality". Multicausality is a view of causality that is more static – or perhaps wholistic- than notions we are standardly familiar with in philosophy. A causal mechanism is

"surrogate relationships" between various social and biological "connotations" of self-identified race. Collins urges "To unravel the real causes, research into health disparities must move beyond weakly correlated variables, such as self-identified race or ethnicity, towards an understanding of the more proximate environmental and genetic factors"(Figure 1, S14).

Collins adds onto the NHGRI's "Vision for the Future of Genomics Research" imperatives for further research identified at Howard (S14):

1. rather than race or ethnicity, which are variables correlated with health outcomes, identify **root causes** of health and disease
2. carry out large-scale studies in multiple populations where **both** genetic and environmental information is collected so that accurate risk factors are determined,
3. carry out long-term, longitudinal prospective cohort studies as well as case-control studies in order to **validate** quantitative conclusions about the importance of genes and environment in health outcomes,
4. support efforts such as the International Human Haplotype Map Project that attempt to define the nature of human variation across the world and use them as resources rather than panaceas,
5. carry out more *anthropological, sociological and psychological research* into how individuals and groups conceive and internalize concepts of race and ethnicity,
6. assess *the way scientists use the concepts of race and ethnicity* and try to remedy cases where the word is misused or used in a counterproductive manner
7. educate researchers, health professionals and the general public with *clear, scientifically accurate messages about the relations between race, ethnicity, genetics and race*.

The last three of these goals invite direct contribution from philosophers, though a method of identifying root causes (i) would also benefit from some philosophical scrutiny.

A2.3 Keita S, Kittles RA, Royal CDM, Bonney GE, Furbert-Harris P, Dunston GM & Rotimi CN, "Conceptualizing human variation"

The piece begins with a section titled "Race: semantics and confusion". Authors distinguish two views on the use of the term 'race': 1. 'races' are real (i.e. there is empirical evidence for the existence of biological 'racial' differences), 2. 'races' are socially constructed (i.e. human agency has had a more important role in creating these distinctions between people) (S17)[197]. It is added that biologists too disagree about what 'race' means and about whether it is an appropriate way of cashing out human infraspecific variation.

The authors think disagreements indicate a problem with semantics: "'Race' is not being defined or used consistently; its referents are varied and shift depending on context. The term is

---

represented as a "causal pie" with pieces of the pie being named "causes" for the disease effect [necessary but not individually sufficient causes (like Mackie INUS conditions)]. It seems that multicausality notices causation at a more "superficial" phenomenal level conceiving causes as always interacting (in possibly ordered ways) at a level perhaps too intricate to map out. It may be a way to pussyfoot around causality but it is an interesting one.

[197] The source referenced here is a piece by philosopher Robin Andreasen in Philosophy of Science, Andreasen (2000).

often used colloquially to refer to a range of human groupings. Religious, cultural, social, national, ethnic, linguistic, genetic, geographical and anatomical groups have been and sometimes still are called 'races'" (S17)[198]. "We argue that the correct use of the term 'race' is the most current taxonomic one, because it has been formalized"(S17) and add that "at this time" the legitimate taxonomic concept of 'race' does not apply to humans (S19).

The 'correct' taxonomic concept of race is equivalent to 'subspecies', authors say. They note some taxonomists reserve the term 'race' for local breeding populations, with 'subspecies' applying to populations with the same or similar defining traits. The definition of subspecies they favor is that of Avise and Ball:

> "Subspecies are groups of actually or potentially interbreeding populations phylogenetically distinguishable from, but reproductively compatible with other such groups. Importantly, the evidence for phylogenetic distinctions must normally come from the concordant distributions of multiple, independent, genetically based traits" (S18).

They explain the definition is favored because it emphasizes phylogenetics and so "evolutionary relationships rather than arbitrarily determined somatic traits"[199].

Note here the contrast and similarities between this approach and Sally Haslanger's. Haslanger also defines a correct, "target" notion of race but does so along the lines of social oppression, and claims this is not a biological but rather a social kind. The concept of race defined here as applied to humans would be a biorace race notion. It takes race to be a biological kind and it's more similar out of the ones already mentioned to Hardimon's biological race concept and the cladistic race concept proposed by Andreasen. It understands race as a classification sorting populations in terms of genetically based biological traits. But do these authors disagree? No! Not as radically as one would expect.

Human geographical variation does not reach a threshold important enough to merit the distinction of subspecies, it is argued. "'Race' is 'socially constructed' when the word is incorrectly used as the covering term for social or demographic groups. Broadly designated groups, such as 'Hispanic' or 'European American' **do not meet classical or phylogenetic criteria for subspecies** or the criterion for a breeding population" (S18). They explain that 'Negro' and 'Caucasian' were terms of early European science used in politics and law and converted into social identities. But whenever 'race' was assigned as a proportion of ancestor race, evolutionary mechanisms were totally disregarded[200]. "The entities resulting from these political machinations have nothing to do with the substructuring of the species by evolutionary mechanisms" (S18). They add the much quoted facts: 1. individuals from one human 'race' may be overall more similar genetically to individuals from another 'race' than to an individual from their own, 2. individuals with the same morphology do not cluster with each other by lineage and 3. a given lineage include only individuals with the same 'racial type' (S18-19).

Should "race" categories be used to approximate "medically interesting humans genetic variation"?

The authors argue *even if there are no human races, 'racial' classifications can and should still be used in human genetic research*. They cite two competing hypotheses as to how genetics contributes to complex disease:

0. The common disease- common variant (CDCV) hypothesis posits "much of the genetic variation contributing to disease is old and shared by most human populations", so health

[198] Nancy Stepan is referenced as source here.
[199] Michael Hardimon's suggestion for what he calls the "biological race concept" seems to agree with the older views as Hardimon's concept is modeled after Mayr's biological species concept.
[200] In contrast perhaps to the ancestry-tracing services provided by Professor Kittles, perhaps. See also the BBC documentary *Motherland*

disparities among population groups are largely due to common exposure to environmental insults (S19).

1. The multiple rare variants (MRV) hypothesis posits "a substantial proportion of genetic polymorphisms will be rare and will probably be specific to groups that experienced similar evolutionary forces of selection or drift" (S19).

The authors point out that both of these hypotheses may be right, depending on the phenotype in question.

Why should racial categories still structure biomedical research? In the case that the MRV hypothesis is true it will turn out that no particular map of polymorphic markers will be sufficient to explain the interplay of forces between genetic variants and genetic factors in complex disease etiology. Thus, they reason, it would be premature to discard racial group classifications because it would be premature to discard any group identifiers that would enable us to see group patterns in disparities. For example, the finding that a demographic group called 'African American' has a higher prevalence of prostate cancer is not to be denied, even if there is no African American race. They go further to propose study populations' history should be recorded more precisely; well-defined local populations within demographic groups should be studied if gene-environment interactions are to be understood.

Instead of discarding 'racial' classifications, on the basis that they are socially-constructed and so biologically meaningless these authors urge we should keep studying race/ethnicity classes and look for classifications that capture patterns of disease disparities, within these broad categories. Perhaps some future classification will correspond more closely to taxonomic 'races' but what takes priority is describing the phenomenon at issue: disease disparities across human groups.

A2.4 Sarah K Tate & David B Goldstein, "Will tomorrow's medicines work for everyone?"

This article claims pharmacogenetics –the clinical use of genetics to identify genetic influences on drug response- could contribute to future health disparities. Authors say in most cases health disparities depend on environmental not genetic factors but this could change in the future as genetics could make medicine exclusive and do so along ethnic or racial lines (S34).

Tate and Goldstein identify three domains of prime risk:

1. Pharmacogenetic diagnostics –individual genotypes could be used to select one's medicines,
2. Disease classification –genetic features of a disease can be used to guide therapy choice,
3. Pipeline pharmacogenetics- genetics can be used for the evaluation of new chemical entities.

The authors admit genetic variation is small among humans and most medicines will tend have similar effects across populations. Still discovering medicine that is maximally effective for its target population is a live possibility and these risks must be tackled.

'Table 1 Examples of drugs reported to have different response in different racial or ethnic groups' lists 22 examples of medicines that claim racial differences in either efficacy or safety (S36-S37).[201] According to Tate and Goldstein few claims have good supporting evidence but 8% of new drugs have claimed racial differences in effectiveness.

Genetic epidemiology methods include comparisons between populations originating from different geographical locations, or studies of admixed populations, but the authors note the best way to evaluate the contribution of genetics to the disparities would be to identify the causes of

[201] At the time of writing their article they say 29 drugs had made such claims.

variable drug response and locate differences in causes between different groups[202]. What limits the conclusiveness of these results is that there have not been enough studies of drug response, nor enough studies with the same study design. Further, most studies compare European American and African American populations, not transcending the bounds of the American cultural context. Finally, locating the cause of a differential drug response might involve complex causal mechanisms as shown by the cited example of a differential response between African Americans and European Americans to beta blockers, ACE inhibitors and angiotensin-receptor blockers (S34). Even when real differences are discovered attributing them to the genetics as opposed to the environmental correlates of race or ethnicity is dubious.

The authors distinguish genetic factors specific to the drug used from those that have to do with the nature of the condition itself. Pharmacokinetic and pharmacodynamic factors specific to a drug are gene variants that causally influence response to treatment and they often have different frequencies among different racial or ethnic groups: "Of 42 genetic variants that have been associated with drug response in two or more studies, more than two-thirds have significant differences in frequency between people of European ancestry and people of African ancestry" (S35). The average frequency difference for all 42 variants is 0.15 and almost a third of the variants show differences of 0.2 or larger (S35).

They claim that contrary to what other researchers say there are genetic variants influencing drug response that differ among different racial groups and even if racially variable therapeutic responses have *no fundamental genetic causes*, race-specific therapy should not be ruled out as invalid since its effects don't presuppose such genetic differences; "In the case of BiDil, it is not currently know whether it works differently in African Americans and European Americans because of genetics, environment or both" (S37).

How are races thought of here? There is no explicit attempt to define the term 'race'. Rather the term is picked up as used in the biomedical literature quoted. Both genetic, biological differences and environmental variation is taken to fall under these categories. And so races are thought of as either or both bioraces and socioraces.

Should "race" categories be used to approximate "medically interesting human genetic variation"?

The authors conclude:
Genetic differences between groups are graded rather than dichotomous, so instead of considering race or ethnicity as a proxy for a genetic factor it is better –when possible- to identify the factor itself. But even when the genetic structure of a population is taken into consideration, race or ethnicity might *still be appropriate to consider* as it might work as a proxy for environmental correlations (S37).
Though individual genotype is more informative than racial or ethnic labels (for genetic effects) there are cases where race and ethnicity may be "useful biological proxies for the underlying genetic variation" (S39). "There are many examples of variants that are known to influence drug response and that differ substantially among racial or ethnic groups (…) This source of genetic variation can and should be represented during our period of ignorance of the underlying causal factors" (S39).

These conclusions illustrate how "race" can get its foot in the door of genetics research without being endorsed as "biological" or "genetic". Race and ethnicity are acknowledged as rough classifications containing poor quality genetic information but found to be useful proxies for either or both 1. environmental and 2. genetic variation affecting health outcomes.

[202] The method proposed is a method for causal enquiry called by John Stuart Mill the method of difference.

And notice that Tate and Goldstein are also concerned with the "fight against injustice" (Haslanger 2000, 36). They caution race and ethnicity categories should not be used to exclude people from the best available medical treatment. They warn of three ways in which pharmacogenetics may play into *racially* stratified drug therapy:
Pharmacogenetic diagnostics could by genetic testing prescribe the most appropriate drug to the patient by minimizing the probability that the patient has an adverse drug response (ADR)[203].
Tests might perform differently among racial groups because of a) underlying differences in physiology, b) the genetic markers used by the test are only proxies for causal variants and have considerably different predictive properties among different racial or ethnic groups.

Locating the genetic bases of common diseases will enable their classification into subclasses with similar phenotypes but different genetic bases enabling the development of drugs for specific subclasses of the disease. Differentiation in the response of racial populations to diseases has in some cases been found to coincide with race-specific differences in the genetics of the disease. And even if the variant associated with a disease is found in many ethnic groups, its effects may differ. The authors caution the HapMap Project, aiming to describe common patterns of human genetic variation, does not include Native Americans or pacific Islanders in its target populations, thereby inhibiting the development of race specific treatments for these populations (S39).

Finally, developing new therapies using pharmacogenetics will enable a (cost-effective) shift in the old model of drug development from designing large expensive phase 3 trials to carrying out smaller, less expensive trials using genetics. This will result in the exclusion of individuals with unfavorable genetic profiles from clinical trials and could create "orphan genotypes" for which no available treatment is designed (S39).

Tate and Goldstein don't propose better concepts for these purposes. It is up to philosophers to try and articulate what concept goes with these envisioned uses made of 'race', and link these back to the critical analytical discourse.

A2.5 Mountain JL, Risch N, "Assessing genetic contributions to phenotypic difference among 'racial' and 'ethnic' groups"

This paper asks whether recently inferred correspondences between clusters inferred from multilocus genetic data and groups defined by self-identified race and ethnicity (SIRE) are solid enough a foundation to imply between-group phenotypic variation has a genetic basis. The thesis of this paper is that there is no foundation for such an assumption but that 'racial' and 'ethnic' categories remain salient.

After introducing their topic, the authors discuss three approaches for assessing the genetic contribution to between and within-group human variation (Table 1, S49):
Family studies: familial correlations obtained from the study of twins in comparison to siblings or adoptees,
DNA analysis: examining specific DNA variations and assessing between and within group associated risks,
Admixture analysis: estimating correlations between the degree of admixture in a newly admixed population and trait values.
The relative benefits and weaknesses of these methods are discussed.

[203] Tradeoffs between positive reactions and adverse ones are not taken to complicate the issue, as perhaps they should.

> The method of familial correlations can provide estimates of heritability within populations, though gene-environment interactions confound clear conclusions about a genetic explanation for trait differences between groups.
> DNA analysis can give an estimate of genetic contribution to between-group differences when this is due to specific genes. But to minimize genetic and environmental interactions such results are derived and apply only to particular populations and are not valid global genomic estimates.
> Admixture analysis can better provide global genomic estimates on the contribution of genetics to between group differences but environmental confounds still apply –except if the admixture is cryptic or the correlation of the admixture and the trait is at a single or a small number of locations in the genome (S49-50).

Mountain and Risch say that assessing genetic contributions to between-group differences depends on the trait under study. A table examines the genetic basis of five categories of traits - across which SIRE groups could be compared (Table 2, S50): 1. DNA-level traits i.e. DNA sequence variation, including SNPs and STRs, 2. Mendelian traits, 3. Physical traits, 4. Complex diseases and 5. Behavioral traits. They explain:

1. With respect to DNA-level traits, our understanding is quite consistent with that of authors 30 years ago: the low proportion of genetic variation between groups does not exclude the possibility of the existence of some genetic contribution to between group difference[204].
2. Over 1,000 Mendelian traits have been mapped to specific chromosomal regions in the last 20 years. Mutations have been found to be rare and to correspond to geographical regions or populations but markers known to be associated with different disease incidence between groups are few.
3. Variations in physical traits that are used to classify people according to race, such as skin pigmentation, hair form or facial features, seem to be determined by more than one gene and may not correspond to any considerably different allele frequencies at one or a few genes.
4. Complex diseases such as inflammatory bowel disease and neurodegenerative diseases have been summarily linked to specific genes, but complex traits such as diabetes and blood pressure are still to be linked to specific genetic regions.
5. Behavioral traits such as cognitive ability, temperament and athletic ability have not been conclusively linked to any specific genes.

The terms “race” and “ethnicity” are not used through this discussion in preference of the term “group”, meaning “self-identified race and ethnicity” groups or SIRE. This is a more accurate term than race insofar as it specifies how reference is fixed (by self-identification) and refers specifically to the populations under study as opposed to demographic races. These groups are understood as biologically and genetically interesting. But the interest these groups warrant doesn’t, according to the authors confirm that there is a genetic basis for phenotypic variation between common racial groups –which includes physical traits but also disease phenotype-. Differences in DNA-level traits between SIRE groups confirms the “existence of the possibility” that there is a genetic basis for these differences (point a. above). This makes use of an understanding of these groups as biological as opposed to social kinds suggesting a “bioSIRE”, so to speak, notion is at work here.

Should “race” categories be used to approximate “medically interesting human genetic variation”?

[204] Note that this emphasis highlights between group difference as opposed to similarity, in stark contrast to how Lewontin’s ’72 results are usually reported, i.e. to emphasize similarity between individuals. What is significant now, in the “post-genomic era”, is population difference as opposed to similarity.

The authors conclude that standards should be put in place for statements linking race to genetics, while racial categories should still be used as proxies for a wide range of genetic and non-genetic factors contributing to health disparities.

A2.6 Charles N Rotimi[205], "Are medical and nonmedical uses of large-scale genomic markers conflating genetics and 'race'?"

The main emphasis of the piece is on the international HapMap project and how this could be a substantial tool in advancing the field of genetic epidemiology. The discussion of the HapMap project is brought to bear on questions regarding the design of racial drugs such as BiDil in the last section of the piece.

The discussion starts by positing that if the common-disease common-variant (CDCV) hypothesis holds, i.e. genetic variants contributing to complex disease are relatively common across human populations, the HapMap project[206] could contribute to genetic epidemiology by giving us a comprehensive catalog of common variants. The HapMap project would not have the resolution required to capture variants responsible for disease if rare variants or interactions between rare variants are responsible for complex disease as populations sampled by the HapMap are too few to give us the requisite detail in mapping human genetic variation. Further problems with the HapMap project would arise from the use of its results in an unethical manner in an attempt to validate "old notions of 'race'" thought to be associated with specific behavioral characteristics [S44].

Still, Rotimi concludes that this risk taken by researchers designing the HapMap project (Yoruba, Han Chinese, Japanese and Americans of northern and western European descent) is justified given: 1. evidence that there are differences in haplotype structure and frequency across populations, 2. ethnic and population information would enable us to choose the most efficient sets of SNPs for association studies, 3. removing population labels would create a false sense of security since the available information could be easily reconstructed giving publicly available information, 4. including such ethnic labels would allow "researchers and ethicists to provide better context for interpreting the biological importance of genetic findings that are associated with particular population identities" [S44].

Should "race" categories be used to approximate "medically interesting genetic variation" that occurs between humans?

The author answers in the negative. Instead of relying on socially defined proxies of genetic relatedness like 'race' Rotimi urges "We should allow the genome to teach us the extent of our evolutionary history without abbreviating it with preconceived notions of population boundaries and social identities" [S44]. Rotimi argues that what follows from genomic information about human population structure is neither a refinement nor a confirmation of 'race' classifications. He posits two reasons that are of different type. First, he says individuals in genetics studies can have membership in more than one biogeographical clusters. Second, the borders of such clusters are not distinct and remain influenced by sampling strategies. The first reason speaks to the content of the hypotheses at issue and the second to the method used to confirm it. He concludes that

---

[205]Charles Rotimi will direct the new NIH Intramural Center for Genomics and Health Disparities (NICGHD), as announced on March 17 2008; cf. http://www.nih.gov/news/health/mar2008/nhgri-17.htm [Last accessed 10/09/08] http://www.nytimes.com/2008/03/18/washington/18nih.html?fta=y [Last accessed 10/09/08]

[206] The HapMap project is designed to map out regions in the human genome where individuals are most likely to differ. See http://www.hapmap.org/whatishapmap.html for a description. [Last Accessed on Dec 10 2008] The International HapMap Project. *Nature* 426, 789-796 (2003) and International HapMap Consortium. Integrating ethics and science in the International HapMap Project. *Nat. Rev. Genet.* 5, 467-475 (2004).

genetics cannot be consistently used to define racial groups: “the process of using genetics to define race is like slicing soup: ‘you can cut wherever you want, but the soup stays mixed’” [S44].

As a consequence of the continuity of genetic variation, and its discordance with race, Rotimi argues that drug development *should not* target commonly defined socio-political groups. Using the example of BiDil, he argues that the “seeming success of an ‘ethnic drug’” only contributes to and highlights our confusion about the relation between race and biology. Rotimi claims ‘race’ is not an adequate proxy for the populations likely to benefit from a specific drug and the full curative potential of the drug might not be reached if associated with a race or ethnicity. One of the justifications that Rotimi offers for the marketing of ‘ethnic’ drugs is that “Race-based hypotheses in biomedical research sell”.

Conclusions

The majority of these articles argue for retaining the use of race and ethnicity categories in genetics research, in lieu of more accurate but harder to measure genotype-specific classes. It seems the argument from genetic premises comes down to an argument from ignorance: why not use these categories? We seem to have captured some variation at this fineness of grain so at risk of losing sight of any pattern let us at least, for now, hold these ones in view. However, in many of these articles there is also a vision for the future of genetics research on race and health outcomes. In the words of Francis Collins, “A true understanding of disease risk requires a thorough examination of root causes” and “research must move beyond these weak and imperfect proxy relationships to define the more proximate factors that influence health” (S13).

Noticing the fluidity and complexity of race categories has been according to Rose (2007) (who quotes Reardon (2004)) one of the main trends in how race categories are discussed in genetics. This analysis concurs with these observations on the level of how people talk about race. But talk of fluidity and complexity contradicts the ease involved in measuring “race” as opposed to the difficulty of measuring what precise non-blurry (whatever these are) populations that race is supposed to approximate: in practice measuring people’s “race” involves no fluid mechanics, only counting discrete numbers of boxes checked, even though the entity itself looks blurry.

I exposed this discussion in this form because what I’m interested in the DIAGNOSIS of this thesis is how we deal with “surrogate relationships” in science and what I care to offer is an ontological frame for how to understand what a surrogate may be. I posit in theory that scientific “innovation” involves using available stuff in new ways and as such is intrinsically about stand-ins. These new “discovered” beings/ techniques/… are in the best case scenario only “recovered”: found as new and used to do stuff they a. could always have done –but were never given the chance to or, b. can do (now). The choice between a. and b. depends on your metaphysics of powers and capacities and how enduring you think these are.

APPENDIX 3

'Race' Founded in Demography

What got me thinking about the possibility of transfiguring a concept was a close reading of a policy document: the recommendations of an advisory committee regarding U.S. race/ethnicity standards. It is to these documents I turn now.

The purpose of the exposition is three-fold:

- first to investigate these demographic concepts and argue that they are not properly speaking concepts of race but rather founded race concepts; concepts founded in the quasi-scientific context of demography to measure different aspects of the U.S. national population,
- second to show that the term 'race' is found in biomedicine as a non-scientific but already available category, and
- third to argue that the term though explicitly acknowledged as non-scientific becomes is "picked up" as of interest in contexts of biomedical, scientific interests.

"Race" becomes of interest because of two uses that founded concepts in demography can be put to: a) indicating the social status and b) the biological status of an individual. The story starts with a flashback to the time when race and ethnicity standards were first standardized across states.

I begin my analysis with a document written in May 1977.

A3.1 1977 Race and Ethnicity Standards for Federal Statistics

After work starting in 1974, the Office of Management and Budget (OMB) revised proposed "*Race and Color* Designations in Federal Statistics" and issued Statistical Policy Directive No. 15, "*Race and Ethnicity* Standards for Federal Statistics and Administrative Reporting". The categories to be used by all federal agencies collecting and reporting race and ethnicity data were defined to be a minimum set of four "races" and one "ethnicity"[207]:

- **American Indian or Alaska Native**
  A person having *origins* in any of the original peoples of North America and who maintains cultural identification through tribal affiliation or *community* recognition.
- **Asian or Pacific Islander**
  A person having *origins* in: a) any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent or the Pacific Islands. This area includes, for example, China, India, Japan, Korea, the Philippine Islands, and Samoa.
- **Black**
  A person having *origins* in any of the black racial groups of Africa.
- **Hispanic**
  A person of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish *culture or origin, regardless of race*.

[207] See: http://www.whitehouse.gov/OMB/fedreg/directive_15.html, under Appendix 1, Directive No. 15 Race and Ethnic Standards for Federal Statistics and Administrative Reporting. And: Wallman KK, Hodgdon J. "Race and ethnic standards for federal statistics and administrative reporting". *Statistical Reporter*, July 1977 (no. 77-10):450-54.

- **White**
  A person having *origins* in any of the original peoples of Europe, North Africa or the Middle East.

Why was this type of data important to collect in the first place?

A3.2 "Race" as of Interest to Politics

Those setting up the standards give an answer; but it seems quite uninformative. The defining 1977 document puts the definition and standardization of these categories down to a need for collecting "compatible, nonduplicated, exchangeable racial and ethnic data"[208].

This seems as vacuous as replying to the question "Why did you name your daughter '*Bizarre*'?" by

–I needed to call to her.

The statement P or Not P is trivially true, so specifying the complement of P should give some idea about what P is. So look at the disclaimers.

What is it said was not said?

> "These classifications should not be interpreted as being scientific or anthropological in nature, nor should they be viewed as determinants of eligibility for participation in any Federal program"[209].

This says categories should be kept safe from racist use, so presumably what the categories were *not* developed for is the political exclusion, insidious stereotyping and essentializing of people along these classes. Indeed, in a later document published by the OMB, the categories are described as having originated "in large measure" in "new responsibilities to enforce civil rights laws"[210]. (It is also noted that the category "Hispanic" was included to satisfy Public Law 94-311, voted in 1976, which required the collection of more social and economic data on persons of Spanish origin or descent. Though why this law was voted in is not specified in the document.)

I point out this discrepancy in the motivations printed and the ones endorsed and twenty years after the "fact" because I am interested in the character of the discrepancy. The divergence between motivations printed and manifest and the ones endorsed and possibly operative already at the time indicates an understanding of these categories that was not an easy one to make relevant to demography at the time as it was later to admit. In 1997 political motivations for setting up these standards were endorsed by the OMB as well as the nature of the categories but twenty years earlier it was a statement of immediate consequences (getting data along these categories) rather than some particular rationale that was used to justify their adoption.

This interest indicates that these categories were not simply decided to describe the phenomenon of "race/ethnicity" on the basis of an ordinary concept of race; but rather to be used as tools for measuring a phenomenon of interest that was properly speaking different than race. This is a first indication that an ordinary concept of race was subject to transfiguration so as to fit the interests of this quasi-scientific demographic context. But all one can infer from the documents is that the categories would be founded in the context of demography as a concept that could track civil rights violations. Looking at later documents makes it easier to see that these categories are founded race concepts.

A3.3 "Race" as of Possible Interest to Biomedicine: Diversity

[208] See: Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting, http://www.whitehouse.gov/OMB/fedreg/directive_15.html, Appendix 1

[209] See: Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting, http://www.whitehouse.gov/OMB/fedreg/directive_15.html, Appendix 1

[210] http://www.whitehouse.gov/OMB/fedreg/directive_15.html, in section A. Background

Given how parsimonious legislative documents may be in what they make manifest, let us examine another document: the "Report to the OMB on the Review of Statistical Policy Directive No. 15". The document sums up recommendations of an *Interagency Committee for the Review of the Racial and Ethnic Standards* (ICRRES) set up in 1994 with the purpose of reviewing and revising the 1997 standards.

As the OMB reports, the revision of race and ethnicity standards was mandated by the "increasing diversity of our Nation's population that has resulted primarily from growth in immigration and in interracial marriages"[211]. So increasing rates in 1. immigration and 2. mixed race marriages put in question the capacity of existing race and ethnicity categories to reflect the makeup of the US population. People immigrating to the US and children born to parents of different races post 1977 were increasingly eluding classification by existing race and ethnicity categories.

Further, two opposing interests are manifestly accepted as those shaping the review process: 1. capturing "increasing diversity" and 2. maintaining continuity with race and ethnicity data already on file. The first interest is an interest in change. It is an interest in developing categories which capture "new" phenomena. The second interest is an interest in continuity. It is an interest in retaining categories which cohere with "old" data.

One can see these interests expressed in five issues identified as central to the process[212]. The questions of capturing increasing diversity while maintaining continuity with existing data are further broken down to the following questions:

1. including a "multiple races" or "multiracial" category,
2. combining questions on race and Hispanic origin,
3. assessing the public's understanding of the concepts of race, ethnicity and ancestry and combining questions about these categories,
4. modifying or replacing current terminology and
5. including new classifications to the minimum basic set.

Further, questions under points 1, 2, 4, 5 make manifest the interest in modifying current categories or introducing new ones and 3 suggests that ultimate modifications depend on comparing the decided classes correspond to a public's understanding of "the concepts of race, ethnicity and ancestry" when asking questions about the categories.

But there is a fixed point in the process: the OMB fixes the "starting point" for the review process in already existing standards and "the recognition of the importance of being able to maintain this historical continuity" (my emphasis)[213]. Any new categories would have to be subsumable in their greatest part under the existing categories to utilize already available data.

These concerns with change and continuity seem endemic to all sorts of modification and are justifiable in their form. But what in particular about their new forms is at issue here? Though concerns with continuity are made explicit in the report what "increasing diversity" is and why it needs to be documented in the census is not. It is only after this is articulated that good standards can be found.

Given standardizations originated in concerns for political equity are we licensed to infer that it is the importance of these categories as proxies for political ("social") inequality that fuels the U.S. government's concern regarding diversity? Is it because it matters for political equity that we now care to count children of "mixed" marriage born per annum in the U.S.?

You can guess I want to say no! It is not *just* the social status of individuals counted under these categories that is said to matter now. But let me first establish that there is a switch. What we expected race and ethnicity categories to do in 1997 seems different *in principle* from what we expected them to do in 1977.

---

[211] http://www.whitehouse.gov/OMB/fedreg/directive_15.html, in section B. Review Process
[212] Ibid, section 1.5
[213] http://www.whitehouse.gov/OMB/fedreg/directive_15.html Supplementary Information, section B.

To do this, reason by contradiction. If we cared for the categories to do now (in 1994) what they were doing then (1977), then why accuse them of lacking a quality now they never had to begin with?

"Accounting for diversity" may have been a problem for the categories in 1997 but it was a problem *already* in 1977: the U.S. population was *already* more diverse in 1977 than the four race and one ethnicity classifications allowed for. Census categories were not federally standardized until 1977 but there are adequate historical records of what would count as racial diversity that partially reflect waves of immigration to the US and date back to the eighteenth and nineteenth centuries[214]. And naively thinking of how Americans refer to their 'origins' one might also posit "ethnicity" categories could describe members of the Irish, Greek or Middle-Eastern U.S. communities already in 1977.

So why wasn't this critique launched at the time?

Diversity, in general, does not seem to have been a concern of census standards of old. From 1790 until 1850 decennial census data used only the general categories "White" and "Negro" with the further specification of "slave" or "free" under the category "Negro". "American Indian" people were not counted if not taxed because the numbers of individuals that were neither white nor negro was reportedly "extremely small"[215]. The 1860 decennial census included for the first time the categories "American Indian" (still excluding those not taxed), Eskimo and Aleut, "Chinese" (in California), Asian and Pacific Islander[216]. The next census in 1870 identified "Japanese" as a separate race and more categories were included in 1910, such as Filipino, Hindu and Korean at a time where significant waves of immigration to the U.S. were documented[217], while the 1930 census used "Mexican" as a race category, which the 1940 census eliminated[218]. There was no "other race" category available until 1940.

Race/ethnicity standards were not manifestly claimed nor manifestly used as natural science categories to document some god-given phenomenon precisely; they were rather manifestly tools for the management of a nation's human, social and political, affairs. They did not aim to document all diversity but rather document the sort of diversity that was of interest: tax-paying ability, or eligibility[219].

What sort of diversity is it then that is now of interest and not captured by the categories?

The review committee reports data that show "Increases in Interracial Marriages and Households and Births to Parents of Different Races"[220] since 1970.

1. It is reported that "interracial unions" increased from 321,000 in 1970 to about a million in 1980 and 1.5 million in 1990 –but no comparative data is given for the increase of the US population between these times and the increase of same-race unions.
2. Census data is again cited for the increase of the number of children in interracial families, which grew from less than half a million in 1970 to about two million in

---

[214] Campbell Gibson and Kay Jung, "Historical Census Statistics On Population Totals By Race, 1790-1990, and By Hispanic Origin, 1970 to 1990, For Large Cities and Other Urban Places in The United States", Population Division, Working Paper No. 76, US Census Bureau, Washington, D.C. 20233, February 2005.

[215] Ibid.

[216] Ibid.

[217] Nayan Shah (2001) gives an account of Pacific Islanders' migration to California at the turn of the century.

[218] The category reportedly failed to count these individuals who either identified as "white" because of an attempt to hide their origin or because they thought "white" applied to them better than "Mexican". Cf. Campbell Gibson and Kay Jung, "Historical Census Statistics On Population Totals By Race, 1790-1990, and By Hispanic Origin, 1970 to 1990, For Large Cities and Other Urban Places in The United States", Population Division, Working Paper No. 76, US Census Bureau, Washington, D.C. 20233, February 2005.

[219] The "one drop" rule similarly simplified the classification of racial births in the U.S. though racial categories for mixed births were available, developed and used in Spanish and Portuguese colonies as early as the 16th and 17th centuries as Mazzolini (2007) reports.

[220] http://www.whitehouse.gov/OMB/fedreg/directive_15.html Appendix 2, Report to the Office of Management and Budget on the Review of Statistical Policy Directive No. 15, Section 3.5.1.1

> 1990. But again no data is given for establishing the relative rates of change for children of same-race families.

This is not then properly shown to be an "increasing" diversity; though it is inherited, biological diversity; that is if we follow common notions of race.

Social scientists Nikolas Rose, Steven Epstein make the claim that *race/ethnicity categories* are interesting because they are found to measure differences in the biological status –not just the social status- *of individuals*. I take that my observations here support just this claim which they argue on different grounds. What stimulates this interest is noticing a pattern, which repeats words of a common, standardized federal vocabulary.

A3.4 Different Interests in Using 'Race'

ICRRES, the *Interagency Committee for the Review of the Racial and Ethnic Standards*, numbered representatives from more than 30 agencies listed as interested in race and ethnicity data. The Department of Health and Human Services numbered the most offices with nine participating offices listed as using race and ethnicity data, while the Department of Justice was the only other agency to have more than two offices participating: it had three[221].

Here "forces" of change and continuity start to get some flesh on their bones (or rather origins and direction). Lower-level operative notions of race are vying for the stamp of the institution's approval.

Section 3.6.1 of the "Report to the OMB on the Review of Statistical Policy Directive No. 15" is titled "Meeting Legislative and Program Needs"[222]. The first paragraph makes explicit reference to federal agencies using data on race and ethnicity for "policy development, program evaluation, and civil rights monitoring and enforcement", i.e. goals much along the lines of the reported original motivation for standardizing the categories. These agencies are said to be "concerned" that allowing people to report more than one race or adding a multiracial category will break continuity with already existing data on which they have to rely to meet their program goals making it hard to follow changing trends in, for example, equal employment opportunities.

This is one intended usage of race/ethnicity concepts: as found social policy targets. It is as concepts appropriate for these programs to take place and carry forth their targets that the agency representatives are considering and evaluating candidateformulations for the new standards.

The next paragraph juxtaposes these worries to those of "other federal agencies that measure and report on various conditions"[223]. These agencies suggest "*the interest* in the reporting of multiracial information reflects a growing phenomenon that will have to be addressed sooner or later" (my emphasis)[224]. The example the review committee gives here is that of "the health field" where "it is important to collect comprehensive data about the racial heritage of individuals" (my emphasis).

The specific example the review committee brings up is telling. They talk about children[225]:

> Studies have indicated that rates of low birth weight, very low birth weight, pre-term delivery, and small-for-gestational-age –key indicators of children's health status –were highest when both parents were Black, followed by rates for children with Black mother/White father, White mother/Black father, and both parents White.

They conclude:

---

[221] http://www.whitehouse.gov/OMB/fedreg/directive_15.html in Appendix 2, Report to the Office of Management and Budget on the Review of Statistical Policy Directive No. 15, section 1.4

[222] http://www.whitehouse.gov/OMB/fedreg/directive_15.html ibid. Section 3.6.1

[223] http://www.whitehouse.gov/OMB/fedreg/directive_15.html Appendix 2, Report to the Office of Management and Budget on the Review of Statistical Policy Directive No. 15, Section 3.6.1

[224] ibid

[225] ibid

> In the context of health research, a Federal standard that permitted the reporting of more than one race could better accommodate efforts to identify individuals at high risk for certain medical conditions. (Emphasis added.)

Here another understanding of what race can help monitor is offered. Race can, in the context of health research, be able to identify individuals' risk for different medical conditions. Continuity is then desired between new standards and categories that can best monitor civil rights; change would be towards categories that can best provide biomedical services. The first categories would be founded in the context of social policy and found social policy targets; the second ones founded in biomedical contexts and found biomedical targets. And recall the starting point that the OMB set to this process, referred to correlatively the importance of maintaining historical continuity with already available data. This is an interest standing as opposed to any of the many ways possible agencies may see it fit to rearticulate the standards, and imposed by higher-level interests in the OMB institution.

A3.5 Concepts Aligned?

So, why is it (now) interesting to capture "increasing diversity"?
We get two answers:

1. First, there is interest in reporting this diversity; *respondents* are interested in *reporting* more than one race,
2. Though social risk is reportedly harder to track when historical continuity is broken, new kinds of risk –the risks of bodily harm and sickness – are reportedly tracked better when the racial status of individuals is more "comprehensively" represented.

There is a possible point not raised in the discussion (among of course millions). But one interest whose absence may be of significance. What is dropped out of the debate is that social inequalities could also have been sensitive to racial "admixtures".

The proximal causal mechanism for racial admixture is a prima facie biological one –it involves the copulation of members of a different race; the distal social conditions that might have brought these two people together remain unarticulated in this frame. But could it not be that tracking this "biological diversity" would be important for social reasons? The department of justice might have argued that children of mixed marriage have differential rates of incarceration depending on their parents' race or ethnicity and stressed the importance of measuring "diversity" on these grounds. But they don't in this document (–which doesn't mean they don't elsewhere).

One may still wonder why at this point. Is it because there really is no indication that such diversity matters for the concerns of this agency? –This is quite likely if as agency representatives suggest the standards used to report such concerns would not cohere with and so not be attentive to this type of diversity.

Or is it perhaps because a founded race concept in the context of legal or social policy practice posits that these are social categories and that in this context the important avenues to check and control for how 'risk' travels are social ones and not prima facie related to the subject's "inherited" diversity? Is it because there are no good bridge principles for accommodating such an understanding of race in the social policy context that such an interest is lacking on their side?

All of the above is very tentative you will appreciate. Further, this is a naïve understanding of social risks; and not one I am attributing as a manifest understanding to agents in this field. But is it operative still?

All these are questions I can sensibly raise, in this frame. It is, in any case, the measured difference in the health outcomes of racially mixed infants that is placed by the discussants themselves at the antithesis of tracking unemployment trends across already defined groups.

The dialectic between

- preserving the categories vs. modifying them

negotiates competing interests in

- maintaining historical continuity vs. accounting for increasing diversity

articulated in practice using tools of
- social science vs. biomedical science

as
- tracking trends in social inequalities vs. tracking trends in health disparities.
- This is phrased as such on the basis of assumed ontological distinctions

between
- social identities vs. biological identities

with
- social capacities/needs vs. biological capacities/needs

and our concern for (and of)
- the past vs. the future.

A3.6 Recap

I made three claims: 1. A *new* demand is manifestly placed upon race/ethnicity standards circa 1997, a demand to track a perceived "increased diversity", 2. The argument for tracking increased diversity gets mileage from the fact that already existing categories are used to measure the distribution of health outcomes in racially mixed groups, 3. The debate between preserving current categories and using new ones is a relief of conflicting and correlative understandings of 'race': that of race as tracking political outcomes and race as tracking health outcomes.

This analysis is offered here as a springboard for further questions[226].

The structure of this debate points to assumptions about what kinds of phenomena race categories could or should be used to track. It is the ordinary notion of race which, it seems, allows this interpretation to go through. But it is no longer just races that we're talking about when talking about these standards' usage; in these contexts. These categories get invested with concerns particular to agencies' missions and capabilities; they get interpreted as the races that can be made use of most efficiently in this context. These are founded race notions.

A3.7 Current Race/Ethnicity Standards

Race/ethnicity categories currently used are to be a minimum of the following five (changes to the 1977 categories are underlined):[227]

1. **American Indian or Alaska Native**
   A person having *origins* in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or *community* attachment.
2. **Asian or Pacific Islander**
   A person having *origins* in: a) any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam or b) a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
3. **Black or African American**
   A person having *origins* in any of the black racial groups of Africa.

[226] The fact that these two juxtaposed interests are linked and discussed as competing and completing each other is evidence of what Rose calls the somatic individuation of biopolitical paradigms

[227] Found on-line at the OMB website; OMB's Statistical Policy Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting: http://www.whitehouse.gov/OMB/fedreg/directive_15.html --*emphasis* added

4. **Hispanic <u>or Latino</u>**
   A person of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish *culture or origin, regardless of race*.
5. **White**
   A person having *origins* in any of the original peoples of Europe, North Africa, or the Middle East.

Changes were made in every category except white. Alternative names were provided for 'Black' and 'Hispanic', on the basis that the terms 'African-American' and 'Latino' were acceptable to the communities and did not confuse respondents. Hispanic was retained as a separate ethnicity category, though the advisory committee noted considerate overlap between the concepts of race and ethnicity, especially for Latin American respondents. Referring to the work of Clara Rodriguez (1994) and Crews and Bindon (1991), Latin American respondents were described as holding a "more social view of race" which tends to include morphological and social characteristics "besides color" and could result in different taxonomical systems[228].

Given the need for historical continuity in the data, it is not surprising that the 1997 revisions did not introduce any new categories nor radically modify the determinants of existing categories. Instead of including a 'multiracial' category, the instructions for identification were revised to allow identification with more than one of the existing categories –a category of 'more than one race' would be included when the data was aggregated.

When asking questions separately on race and ethnicity, the committee advised asking about ethnicity (Hispanic/Non-Hispanic) first. It noted (again) the inability of respondents to clearly distinguish between the concepts of race, ethnicity and ancestry and the tendency of Hispanic populations to view this identity as a "racial" one. Finally, in the noted absence of federal pressure to track candidate ethnicity categories such as 'Arab or Middle Eastern' and 'Cape Verdean', no new ethnicity categories were introduced.

A3.8 Operative Race Concepts

It is not surprising that respondents had trouble distinguishing between the notions of race, ethnicity and ancestry. These notions are tightly interwoven and the definitions stated above do little to disentangle them.

To be fair, it is not the purpose of OMB race and ethnicity categories to articulate the concepts of race and ethnicity. This is philosophical turf. But assuming that there are identifiable concepts of race and ethnicity at work here –and not just a random labeling of these groups as "race" and "ethnicity" classes- it's worth examining whether the definitions would map back onto what we may hope would be the principled conceptual grounds on which we distinguish race from ethnicity.

Let us see what structure is inherent in the concepts of these standards themselves. I will call these 'OMB race' and 'OMB ethnicity' for clarity.

Compare the definitions of the four OMB races to the definition of the one OMB ethnicity. First, each and every definition makes reference to "origins" but the definitions differ on a. what these origins are origins in and b. whether these origins are necessary for identification. All four OMB races are defined in terms of a person's *origins in people* from particular geographic locales, rather than origins in those locales. The mention of "peoples" (or the circular reference to "racial groups" under "Black or African American") suggests that one's lineage decides one's OMB race -and reversely that OMB race is an indication of one's lineage. And though identification with the category "American Indian or Alaska Native" presupposes community affiliation, the only necessary condition common across these OMB race categories is origin in some group of people. So lineage decides OMB race.

---

[228] Ibid, under section 4.2 "Concepts of Race and Ethnicity"

On the other hand, the definition of "Hispanic or Latino" which is not an OMB race makes reference to origins *or* a culture specific to particular geographic regions. Coming from a particular place, or sharing the culture associated with a particular place, "regardless of race" is what decides OMB ethnicity. Contrary to racial identification, ethnic identification could be a matter solely of *cultural* identification. So, lineage decides OMB race; culture or lineage decide OMB ethnicity.

Where does this leave us conceptually?

OMB ethnicity is defined by reference to origins in a place or culture and OMB race is defined by origins in people. Lineage decides OMB race and need not decide OMB ethnicity. Culture cannot decide OMB race but can decide OMB ethnicity. It seems that the distinction between OMB ethnicity and OMB race relies on a distinction between environment/community and lineage.

How is 'OMB race' related to Hardimon's Ordinary Concept of Race?

Recall here (HOCR) specified as follows (Section 1.2):

> (HOCR) The *ordinary* concept of race is our common sense concept of race that contains the logical core of all race concepts. This concept distinguishes human groups on the basis of three criteria contained in the logical core of the concept :
>
> HLC (1) visible *physical features* of the relevant kind
> HLC (2) common *ancestry*
> HLC (3) distinctive *geographic origin*

And recall that the first criterion describes 'race' as a classification that sorts people into groups according to the way they look –in particular according to skin-color, facial features, hair texture, etc. The second criterion captures the idea that 'race' is inherited: "A race is a lineage, a line" says Hardimon (445). The third criterion captures that the origins of these groups are associated with different geographical -usually continental- regions, such as Africa, Eurasia, East Asia, America and the Pacific Islands.

We see that a principle in the logical core of the ordinary notion of race, namely HLC2, ancestry, is preserved in the definitions of the standards. Race is decided on the basis of lineage but ethnicity need not be so decided.

Lineage is not in the "core" shall we say of 'OMB ethnicity' but it is in the core of 'OMB race'. So these standards' definitions seem to accord with that demand placed by HLC(2) placed on race concept. But there is more information here than in Hardimon's HLC2.

By what criteria are we to pick relevant places and relevant lineages and by what degrees of error should we refine our judgments? Why does coming from Mexico decide your OMB ethnicity not your OMB race? Why does having French ancestors decide your OMB race and not your OMB ethnicity?

What is implicit in these definitions is the importance not of place but of time; the importance not of geography but of history. Ancient geographical origin and recent geographical origin, ancient ancestry and recent ancestry make a difference as to what OMB race you should identify with. A time frame is implicitly invoked when reference is made to "origins in *the original* peoples of" a region, rather than "origins in peoples of" a region simply. Indeed, every OMB racial category definition except "black" refers to origins in *original* peoples, with the double use of origins in originals giving here a sense of ancient time.

An observation about the importance of the geographical origins of the human groups distinguished would stands in correspondence to Hardimon's HLC(3). Origins in particular continental regions are what we expect to matter for identifying 'race'. And notice that in these definitions we find in-built the notion of founder populations; of original peoples, coming from these regions. As Hardimon notes the "purity" of any such populations is not a part of the logical core of the notion of race. And here too, we don't see any assumptions about how many ancestors one may have coming from what regions; we don's read a person "all" of whose ancestors originate in the original peoples of region X. Depending on whether these founder populations and lines are taken to be "pure" or not, etc, these would be assumptions that the racialist race

notion (HRCR) or the populationist race notion (HPCR) would specify (Section 1.2). But there is some divergence here between the continental origins of races and OMB races.

Recall, the definitions:

1. **American Indian or Alaska Native**
   A person having *origins* in any of the original peoples of North and South America (including Central America), and who maintains tribal affiliation or *community* attachment.
2. **Asian or Pacific Islander**
   A person having *origins* in: a) any of the original peoples of the Far East, Southeast Asia, or the Indian subcontinent including, for example, Cambodia, China, India, Japan, Korea, Malaysia, Pakistan, the Philippine Islands, Thailand, and Vietnam or b) a person having origins in any of the original peoples of Hawaii, Guam, Samoa, or other Pacific Islands.
3. **Black or African American**
   A person having *origins* in any of the black racial groups of Africa.
4. **Hispanic or Latino**
   A person of Mexican, Puerto Rican, Cuban, Central or South American, or other Spanish *culture or origin, regardless of race*.
5. **White**
   A person having *origins* in any of the original peoples of Europe, North Africa, or the Middle East

These explicitly make reference to continents where races are ordinarily taken to originate from: America; Asia and the Pacific Islands; Africa and Europe. But they are more mixed up. We see included under 'Asia' regions of the Indian subcontinent which by ordinary notions of race would specify origins of a white race. This labeling practice is common across the U.S. and the U.K. and possibly to do with how data and what data it is important to standardize and collect and founded race notions in the historical contexts of colonial empires; it runs counter to a mandate that an ordinary notion of 'race' would impose.

But I mentioned one exception. The exception is the definition of "black" which is unashamedly circular: a person is black or African-American if originating in the "black racial groups of Africa". But even in this case time matters and arguably in more than a conceptual manner here.

The inability of two reviewing committees to identify this definition as circular suggests that either "black" was not considered an ambiguous term –and hence did not need further explanation- or that further explanation was not possible. Of course there is a third option. "Black" in the definition of "black or African American" refers to the color "black". This reading seems idiotic if we take it literally as no person is black. If we take it as a phenomenological judgment then the reference to black color here is dated, loaded and still, mistaken. It should have been easy to replace "*black racial groups* of Africa" by "the *original peoples* of sub-Saharan Africa" in sync with what the other definitions are doing so we must wonder. Why has black race remained its own *definans*? I will explore two options: one is that the notion of "original peoples" is taken not to apply to this group.

Being an autochthon, someone who is "born of" a land and toils this land, rather than a nomad has been a distinction of normative import since ancient times. This distinction was revisited by debates between monogenetic and polygenetic accounts of human evolution in the 18th and 19th century. Monogenetic accounts of human natural history posited the presence of one common human ancestral group and were historically resisted in favor of polygenetic accounts which spoke of multiple human ancestral populations[229]. One reason for preferring polygenetic accounts of human evolution was that sharing ancestors with some "inferior" or "degenerate" race was insuperable by many racialist ideologies. But there are political and legal undertones

[229] Cf. Bernasconi & Lott eds (2003), Introduction, viii

this choice. The debate between monogenesis and polygenesis questioned a people's relation to the land they occupied: Is land given to people by natural or divine law or is it a matter of human choice and struggle? Does land belong to its indigenous people or is it up for grabs? These political questions hover over prima facie scientific ones. Questions of racial equality and questions of human justice are strongly coupled historically as well as conceptually[230]. Groups of outsiders do not need to look and speak very differently before they are typecast as "barbarians".

Given this context, one wonders whether reference to origins in "the original peoples" in these race definitions performs the same normative work as the notion of "autochthony" could. Granted, the United States boasts a host of non-autochthones. People of white origin are nomads to the United States. But their presence follows various narratives like those of "choice" of "pioneering adventure" and so on, in which the New World is unashamedly new and virgin, ready "to be had" by a people who are already experienced, cultured, originating in originals – even if now nomads[231]. If so, one should further wonder why the only group denied such a past, by definition, is the one of "black or African-American". The historical fact that black people were sold into the United States to toil the land of others would not negate their origins in "original peoples" but it could possibly serve to obscure it.

The second is that the notion of "black" is indeed referring to color. But to a notion of color like Haslanger uses in her definition of race; or what Hardimon specifies by HLC (1): visible *physical features* of the relevant kind

This is indeed a criterion of essence if we follow Hardimon to the notion of race, and one not made manifest in these standards definitions. We see no specification of color except in the names of the categories of "Black" and "White". This is interesting. Because a notion that was manifest in the title of the standards to be revised in 1974 ("*Race and Color* Designations in Federal Statistics") is here not manifest, and if not lost, at least de-emphasized as of relevance. Except perhaps in the category "black". Here perhaps it is still color that matters. Which is not to make a claim that the category is understood in racialist ways. As already noted by Root and discussed in (XXX) if racist thinking operates along the lines of color then color information might be the type of information that is relevant to collect, still.

I might be making too much of this discrepancy between the race definitions of races other than black and the one of black. But the discrepancy is glaring to my eyes, so much so that it begged some scrutiny.

All in all, one would say that these particular 'race/ethnicity' notions are not too far off from what we would expect ordinary 'race' and 'ethnicity' notions to describe.

But they are different. They are different in calculated ways.

A3.9 'Races' as Found Individual Variables

The development of current race and ethnicity standards was arguably embedded in a context of conflicting interests about what the categories should do. But what about the first manifest purpose they were meant to serve?

Recall the initial justification given for formulating these standards: The defining 1977 document puts the definition and standardization of these categories down to a need for collecting "compatible, nonduplicated, exchangeable racial and ethnic data"[232].

I called this a seemingly trivial statement earlier. But it is not. Not in the context of demography. There is a non-trivial assumption stated in this seemingly vacuous justification. It is

[230] Mueller-Wille and Rheinberger eds (2007) suggests that the concept of heredity was indeed strongly coupled to notions in the legal realm.

[231] Cornel West (1982) offers an interesting account of the consciousness of a colonizing people who are outcasts, inferior to a glorious, cultured "Europe" while masters to the slaves.

[232] See: Directive No. 15, Race and Ethnic Standards for Federal Statistics and Administrative Reporting, http://www.whitehouse.gov/OMB/fedreg/directive_15.html, Appendix 1

the claim that OMB *race (and OMB ethnicity) IS such that*, data specified by it CAN be collected and aggregated using methods we already employ in census counts.

These are already non-trivial demands placed on what an ordinary notion of race would be expected to do to be properly founded in demography; starting with the first and very obvious one: OMB races are to distinguish between people. Not between groups of people. These would be concepts labeling and classifying individuals, and doing so in ways as precise and clean as possible –ones that can be captured in one or two questions in a questionnaire. [Not discussed in philosophy articles in philosophy journals.]

If we call the race categories "r1" to "r4" and the ethnicity "e1" then the claim made can be broken down to three assumptions:

a) There is a method of identification such that almost any individual in any state of the United States can be identified as belonging to one of the *categories r1, …, r4 and/or e1*,
b) it is possible thereupon to collect and organize data according to the populations the individuals identified or were identified with and
c) the result of this process is an adequately representative set of *race and ethnicity data* on the population of the United States.

I labeled the OMB race/ethnicity categories with shorthand names (r1, r2, r3, r4, e1) in [1] to make their distinction from race and ethnicity characteristics in [3] visible.

Though race/ethnicity categories are not used as mathematical variables in the context of the examined document, the categories are to be deployed as "variables" once it comes to using them in quantitative social and natural scientific research. The step of abstraction from these individuals' characteristics to mathematical variables enables mathematical and in particular statistical reasoning about these categories.

Statistical reasoning is predicated on the assumption that we can coherently speak of populations we have limited epistemic access to on the basis of information that we obtain from samples we consider thereby as *of* these populations. This is a step at which a selection of a concept of what our samples are samples of matters.

Deploying these race/ethnicity categories as variables in empirical research relies on these being "well-defined" characteristics. Can we claim that race/ ethnicity is a well-defined characteristic of the U.S. population? It would be if there was a good map for assigning each individual to one race/ethnicity.

This is the work of social science. How do we map (and ultimately found) the real confusing world to the sorts of things we can deal with in our science? There are methods for ensuring that each individual will self-identify with one of these race categories and/or that ethnicity; asking the right questions and demanding an answer are part of it. But making sure the information we are getting is the needed sort of information is another.

My study of this particular case is not deep enough to say more on the topic. I have argued that OMB races are in some ways similar to HORCs but in other ways distinct. They seem to preserve some of HLC1-3, but not articulate the notions in any specified concept of race. I have proposed this is because the prime function of OMB races is to sort between individuals in systematic ways; and ways of interest to the U.S. context. I have proposed that these are assumptions with import as to what OMB races are. And of impact to the effect that these are not races; but rather found census classes. [Whatever these may be has to be further specified.]

I have not evaluated the aptness of these categories for what they purport to track. Nor the reality of the traits they purport to track. This was a story about the aptness of these categories for the interests of those setting up these standards and how conflicting interests may have shaped the content and methods for deploying these categories.

Appendix 4

A More General RECIPE for Found Science

This formulation captures how I'm thinking about the recursiveness of finding and founding. I speak here of an "entity" and "science" but one may substitute these for other terms depending on the object that is found and the target context this is (to be) founded in.

I propose the following RECIPE (a Route of Entry of a Commonly Interesting but Plebeian Entity) for science.

START: Ordinary or non-scientific entity

1. Finding an entity, as ordinary, in science
   In two types of Context:
   1.1. Context of Spaces of science
      1.1.i Using Finding Tools
      1.1.ii for a Context of Spaces
   OR 1.2. Context of Interests of science
      1.2.i Using Finding Tools
      1.2.ii for a Context of Interests

AND 2. Founding the entity, as scientific, in science
   In two types of Context:
   2.1. Context of Interests of science
      2.1.i (using finding tools) as Founding Tools
      2.1.ii for a Context of Interests
   OR 2.2. Context of Spaces of science
      2.2.i as Founding Tools
      2.2.ii for a Context of Spaces

STOP: The entity can be found as scientific, in science.

This recipe would not tell a story about how scientifically *meritorious* the founded entity is; it would only say why one might think it has any *scientific* merit. Founding an entity in science could be the result of good or bad science practices by whatever standards we care to consider. But it will be good or bad SCIENCE that comes out of the process.

References

Andreasen Robin O. (2000), "Race: Biological Reality or Social Construct?", *Philosophy of Science Supplement*, 67, Proceedings of the 1998 Biennial Meetings of the Philosophy of Science Association, S653-S666

Atkinson Tony, B. Cantillon, E. Marlier and B. Nolan (2002), *Social Indicators : The EU and Social Inclusion*, Oxford University Press

Bachelard Gaston (1984), *The New Scientific Spirit*, Beacon Press

Bhan A, Singh JA, Upshur REG, Singer PA, Daar AS (2007), "Grand Challenges in Global Health: Engaging Civil Society Organizations in Biomedical Research in Developing Countries", *PLoS Medicine*, 4(9), e272

Barbujani Guido, Magagni Arianna, Minch Eric and L. Luca Cavalli-Sforza (1997), "An apportionment of human DNA diversity", *Proceedeings of the National Academies of Science USA*, 94, 4516-4519

Bernasconi Robert and Tommy Lee Lott eds (2000), *The Idea of Race*, Hackett Readings in Philosophy, Hackett Publishing Company

Bowcock A.M., Ruiz-Linares A., Tomfohrde J., Minch E. and Cavalli-Sforza L.L. (1994), "High Resolution of Human Evolutionary Trees with Polymorphic Microsatellites", *Nature*, 368, 455-457

Bowker Geoffrey C. and Susan Leigh Star (2000), *Sorting Things Out: Classification and its Consequences*, MIT Press

Brancati et al. (1996), "Diabetes Mellitus, Race, and Socioeconomic Status, A Population-Based Study", *AEP*, 6, 1, 67-73

Burchard Esteban Gonzalez et al. (2003), "The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice", *New England Journal of Medicine*, 348, 1170-1175

Cartwright Nancy (1983), *How the Laws of Physics Lie*, Oxford: Clarendon Press.
-------- (1999), *The Dappled World*, Cambridge University Press
-------- (2003), "What Makes a Capacity a Disposition?", *CPNSS Causality: Metaphysics and Methods*, series ed. Julian Reiss, Technical Report 10/03
-------- (2006), "Against 'The System'", in Is there Value in Inconsistency?, Cristoph Engel and Lorraine Daston (eds.), Nomos Verlagsgesellschaft, Baden-Baden

Chang Hasok (2004), *Inventing Temperature: Measurement and Scientific Progress*, Oxford University Press

Chang Hasok and Nancy Cartwright (2006), "Measurement", *Routledge Companion to the Philosophy of Science*, Stathis Psillos and Martin Curd (eds), 702-718

Churchland Paul (2007), "Demons Get Out", Interview with Professor Paul Churchland, *COLLAPSE Journal of Philosophical Research and Development*, 2

Cohn et al (1986), "Effect of Vasodilator Therapy on Mortality in Chronic Congestive Heart

Failure. Results of a Veterans Administration Cooperative Study", *New England Journal of Medicine*, 314, 24, 1547–1552

Cooper Richard et al. (2003), "Race and Genomics", *New England Journal of Medicine*, 348, 1166-1170

Criqui Michael et al. (2005), Ethnicity and Peripheral Arterial Disease –The San Diego Population Study, *Circulation*, 112, 2703-2707

Cross Caroline (2006), *Marcel Duchamp*, trans. Vivian Rahberg, Reaktion Books

Danto Arthur (2000), "Life in Fluxus", *The Nation*, September 18, 2000
-------- (1997), *After the End of Art –Contemporary Art and the Pale of History*, Princeton University Press
-------- (1981), *The Transfiguration of the Commonplace*, Harvard University Press

Daar Abdallah S. and Peter A. Singer (2005), "Pharmacogenetics and geographical ancestry: implications for drug development and global health", *Nature Reviews Genetics*, 6, 241-246

Daston, Lorraine and Peter Galison (2007), *Objectivity*, Zone Books

Doppelt Gerald (ms), "Can Value-Commitments Ground Scientific Knowledge?", talk presented at the conference of the Society for Social Studies of Science, October 2007

Dries, Exner et al. (1999), "Racial Differences in the Outcome of Left Ventricular Dysfunction", *New England Journal of Medicine*, 340, 609 -616

Dupre John (1993), *The Disorder of Things: Metaphysical Foundations of the Disunity of Science*, Cambridge Mass and London: Harvard University Press
-------- (1995), *Humans and Other Animals*, Clarendon, Oxford University Press
-------- (ms), "What genes are, and why they have nothing to do with race", talk presented at the LSE in June 2008

Duster Troy (2005), "Race and Reification in Science", *Science*, 307, 5712, 1050 – 1051

Ellison George (2005), "'Population Profiling' and Public Health Risk: When and How Should we Use Race/Ethnicity?", *Critical Public Health*, 15, 1, 65-74

Ellison George et al. (2007), "Racial Categories in Medicine: A Failure of Evidence-Based Practice?", *PLoS Medicine*, 4:9, e287, 1434-1436

Epstein Steven (2007), *Inclusion: The Politics of Difference in Medical Research*, Chicago University Press

Exner, Dries et al. (2001), "Lesser Response to Angiotensin-Converting Enzyme Inhibitor Therapy in Black as compared to White Patients with Left Ventricular Dysfunction", *New England Journal of Medicine*, 344, 1351-1357

Falush Daniel, Matthew Stephens and Jonathan K. Pritchard (2003), "Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies", *Genetics*, 164, 1567-1587

Fava Giovanni A. and Chiara Ruini (2003), "Development and Characteristics of a Well-Being Enhancing Psychotherapeutic Strategy: Well-Being Therapy", *J of Behaviour Therapy and Experimental Psychiatry*, 34, 45-63

Fleck Ludwik (1979), *Genesis and Development of a Scientific Fact*, Chicago, Chicago University Press

Foster Morris W. and Richard R. Sharp (2002), "Race, Ethnicity, and Genomics: Social Classifications as Proxies of Biological Heterogeneny", *Genome Research*, 12, 844-850

Geuss Raymond (1981), *The Idea of a Critical Theory –Habermas and the Frankfurt School*, Cambridge University Press

Gillies Donald (1993), *Philosophy of Science in the Twentieth Century: Four Central Themes*, Blackwell

-------- (2006), "Kuhn on Discovery and the Case of Penicillin", in *Contemporary Perspectives in Philosophy and Methodology of Science*, Wenceslao J. Gonzalez and Jesus Alcolea (eds.), netbiblo, 47-63

Goodman Nelson (1978), *Ways of Worldmaking*, Hackett Publishing Company

-------- (1983), *Fact, Fiction, and Forecast*, Harvard University Press

Hacking Ian (1999), *The Social Construction of What?* , Cambridge Mass: Harvard University Press

-------- (2002), *Historical Ontology*, Harvard University Press

-------- (2005), "Why Race Still Matters", *Daedalus*, Winter, 102-116

-------- (2006), "Genetics Biosocial Groups and the Future of Identity", *Daedalus*, Fall, 81-95

Hardimon Michael (2003), "The Ordinary Concept of Race", *J of Phil*., C.9, 437-455
-------- (ms1), "On the Ontology of Race"
-------- (ms2), "The Idea of a Scientific Concept of Race"

Hardy et al. (2003), "Ethnic Differences and Disease Phenotypes", *Science*, 300, 739, 740

Haslanger Sally (2005a), "What Are We Talking About? The Semantics and Politics of Social Kinds", *Hypatia*, 20, 4, 10-26
-------- (2005b), "You Mixed? Racial Identity Without Racial Biology", Chapter 13 in *Adoption Matters: Philosophical and Feminist Essays*, Sally Haslanger and Charlotte Witt (eds), Cornell University Press, 265-289
-------- (2004), "Future Genders? Future Races?" *The Annual Proceedings of the Center for Philosophic Exchange 2003-2004*, 34, 4-27
-------- (2000), "Gender and Race: (What) Are They? (What) Do We Want Them To Be?", *NOUS*, 34,1, 31-55
-------- (1995), "Ontology and Social Construction", *Philosophical Topics*, 23, 2, 95-125

Heidegger Martin (2002), "The Age of the World-Picture", *Off the Beaten Track*, Cambridge University Press

Hopkins David (2004), *Dada and Surrealism: A Very Short Introduction*, Oxford University Press

Hoyningen-Huene, Paul (1987), "Context of Discover and Context of Justification", *Studies in History and Philosophy of Science*, 18, 4, 501-515.

Jablonka Eva and Marion J. Lamb (2005), Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life, MIT Press

Jasanoff Sheila (1995), *Science at the Bar : Law, Science, and Technology in America*, Harvard University Press

Jones Camara Phyllis (2001), "'Race', Racism, and the Practice of Epidemiology", *American Journal of Epidemiology*, 154, 4, 299-304

Kahn Jonathan (2004), "How a Drug Becomes 'Ethnic': Law, Commerce and the Production of Racial Categories in Medicine", *Yale Journal of Health Policy, Law and Ethics*, IV, 1, 1-46

Kaufman Jay S. and Richard S. Cooper (2001), "Commentary: Considerations for Use of Racial/Ethnic Classification in Etiologic Research", *American Journal of Epidemiology*, 154, 4, 291-298
-------- (2001), "Kaufman and Cooper Respond to "'Race', Racism, and the Practice of Epidemiology"", *American Journal of Epidemiology*, 154, 4, 305-306

Keita S, Kittles RA, Royal CDM, Bonney GE, Furbert-Harris P, Dunston GM & Rotimi CN (2004), "Conceptualizing human variation", *Nature Genetics Supplement*, 36, S17 - S20

King Mary-Claire and Arno G. Motulsky (2002), "Mapping Human History", *Science*, 298, 2342, 2343

Kitcher Philip (2007), "Does `Race' Have a Future?", *Philosophy and Public Affairs*, 35, 4, 293-317
-------- (2002), *Science, Truth, and Democracy*, New York : Oxford University Press
-------- (1999), "Race, Ethnicity, Biology, Culture", in *Key Concepts in Critical Theory: Racism*, Leonard Harris (ed), Humanity Books, Prometheus Books

Krieger Nancy (2001), "Theories for Social Epidemiology in the 21st Century: An Ecosocial Perspective", *International Journal of Epidemiology*, 30, 668-677

Kuhn Thomas S (1996 [1962]), *The Structure of Scientific Revolutions*, The University of Chicago Press

Latour and Woolgar (1979), *Laboratory Life: The Construction of Scientific Facts*, Princeton University Press

Lewontin Richard G (1972), "The Apportionment of Human Diversity", *Evol. Biol.* 6: 381-398.

Longino Helen E. (1990), *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*, Princeton University Press

Mazzolini Renato G. (2007), "Las Castas: Interracial Corssing and Social Structure, 1770-1835" in *Heredity Produced*, ed Staffan Muller-Wille and Hans Jorg Rheinberger, MIT Press, 349-373

McAllister, James W. (1996), *Beauty and Revolution in Science*, Cornell University Press

Mueller-Wille, Staffan with Hans-Joerg Rheinberger eds. (2007), *Heredity Produced: At the Crossroads of Biology, Politics and Culture, 1500-1870*, MIT Press

Pickering Andrew (ed) (1992), "From Science as Knowledge to Science as Practice", in *Science as Practice and Culture*, The University of Chicago Press

Pigliucci Massimo and Jonathan Kaplan (2003), "On the concept of biological race and its applicability to humans", *Philosophy of Science*, 70, 1161-1172

Polanyi Michael (1962), *Personal Knowledge: Towards a Post-Critical Philosophy*, The University of Chicago Press

Preston et al. (1998), "Age-race Subgroup Compared with renin profile as Predictors of Blood Pressure Response to Antihypertensive Therapy", *Journal of the American Medical Association*, 280, 1168

Pritchard JK, Matthew Stephens and Peter Donnely (2000), "Inference of Population Structure Using Multilocus Genotype Data", *Genetics*, 155, 945-959

Reardon Jenny (2004), *Race to the Finish: Identity and Governance in an Age of Genomics*, Princeton University Press

Reiss Julian (2007), *Error in Economics: The Methodology of Evidence-Based Economics*, Routledge

Reiss Julian and Philip Kitcher (2008), "Neglected Diseases and Well-Ordered Science", series ed Damien Fennell, *CPNSS Contingency and Dissent in Science*, Technical Report 06/08

Risch Neil, Esteban Burchard, Elad Ziv and Hua Tang. (2002), "Categorization of humans in biomedical research: genes, race and disease", Genome Biology, 3.7, 2007.1-2007.12

Root Michael (2000), "How We Divide the World", *Philosophy of Science*, 67, Supplement. Proceedings of the 1998 Biennial Meetings of the PSA. Part II: Symposia Papers, S628-S639
-------- (2001), "The Problem of Race in Medicine", *Philosophy of the Social Sciences*, 30, 1, 20-39.
-------- (2003), "The Use of Race is Medicine as a Proxy for Genetic Differences", *Philosophy of Science*, 70, 1173-1183
-------- (ms), "Measurement Error in Racial and Ethnic Statistics"

Rose Nikolas (2007), *The Politics of Life Itself: Biomedicine, Power and Subjectivity in the Twenty-First Century*, Princeton and Oxford: Princeton University Press
-------- (ms), "Race, risk and medicine in the age of 'your own personal genome'", talk presented at Simon Fraser University, Spring 2008.

Rosenberg et al. (2002), "Genetic Structure of Human Populations", *Science*, 298, 2381- 2385
-------- (2005), "Clines, Clusters, and the Effect of Study Design on the Inference of Human Population Structure", *PLoS Genetics*, www.plosgenetics.com, 1, 6, e70, 0660-0671

Rothman Kenneth J and Greenland Sander (2005), "Causation and causal inference in epidemiology", *American Journal of Public Health*, 95, S144-150

Sankar P, Cho MK, Condit CM, et al. (2004), "Genetic research and health disparities", *Journal of the American Medical Association*, 291, 2985-2989

Sankar P and J Kahn (2005), "BiDil: Race Medicine Or Race Marketing?" *Health Affairs*, Web Exclusive, W5-455 - W5-463

Schulman et al. (1999), "The Effect of Race and Sex on Physicians' Recommendations for Cardiac Catheterization", *New England Journal of Medicine*, 340, 618 -626

Schaeffer Jean-Marie (2000), *Art of the Modern Age Philosophy of Art from Kant to Heidegger*, trans. Steven Rendall, New French Thought, Princeton University Press

Shah Nayan (2001), *Contagious Divides –Epidemics and Race in San Francisco's Chinatown*, University of California Press

Star Susan Leigh and James R. Griesemer (1989), "Institutional Ecology, `Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39", *Social Studies of Science* ,19, 387-420

Tang Hua, Tom Quertermous, Beatriz Rodriguez, Sharon L.R. Kardia, Xiaofeng Zhu, Andrew Brown, James S. Pankow, Michael A. Province, Steven C. Hunt, Eric Boerwinkel, Nicholas J. Schork, and Neil J. Risch (2005), "Genetic structure, Self-Identified Race/Ethnicity, and *confounding* in case-control association studies", *American Journal of Human Genetics*,76(2), 268-275

Weber Max (1949),*The Methodology of the Social Sciences*, Free Press
-------- (1978 [1913]) "Value-judgments in Social Science" in *Weber Selections in Translation*, ed. W.G. Runciman, trans. Eric Matthews, Cambridge University Press, 69-98
-------- (1978 [1904]), "The Concept of 'Following a Rule'" in *Weber Selections in Translation*, ed. W.G. Runciman, trans. Eric Matthews, Cambridge University Press, 99-110

West Cornel (1982), *Prophesy Deliverance! An Afro-American revolutionary Christianity*, The Westminister Press

Wilson James F, David B. Goldstein et al. (2001), "Population genetic structure of variable drug response", *Nature Genetics*, 29, 265-269

Wittgenstein Ludwig (2001[1953]), *Philosophical Investigations*, Blackwell

Wood Ghislaine (2007), "Dali's Jewellery" in *Surreal Things –Surrealism and Design* edited by Ghislaine Wood, V&A Publications, 215-225

Worral John (ms), Lecture notes for the course on 'Evidence, Objectivity and Scientific Method', taught at the LSE in the Michelmas term of 2008

Zack Naomi (2002), *Philosophy of Science and Race*, Routledge