Metaphysics and Race

Michael Peter Atkinson

UCL

This thesis is submitted for the degree of Master of Philosophical Studies.

Declaration

I, Michael Peter Atkinson, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, this has been indicated in the thesis.

Abstract

This thesis examines the metaphysics of race. It begins by trying to find an interpretation of the claim that race is socially constructed which makes sense as a position within a substantive metaphysical debate. By identifying the different commitments and controversies in the debate, I argue that the best such interpretation is a constitutive one.

I then consider Barnes's (2020) discussion of the metaphysics of gender, in which she advocates two theses. The negative thesis holds that a successful metaphysics of gender need not line up with ordinary gender terms and beliefs. The positive thesis holds that the metaphysics of gender is concerned with explaining the various phenomena of gender. The expressions required in our ideology and posited categories and entities required in our ontology are those needed to satisfactorily explain gender. In applying this to race, I argue that there is no distinctive explanatory task for metaphysicians of race to engage in. There is no explanatory remainder left once the natural and social sciences have performed their work. The correct metaphysics of race, while concerned with explaining the phenomena of race, is not to be determined by metaphysicians. This is the modified positive thesis. The modified positive thesis fits neatly with the post-Quinean thought that science and metaphysics are continuous.

In Chapter 3, I argue for deflationism regarding the actual metaphysics of race debate. There is an ontological consensus - an agreement over what, empirically, there is - between the participants. What they disagree over is what racial concepts (should) apply to. But, while interesting and useful, this isn't a substantive metaphysical debate. The debate over how to

explain race, on the other hand, is substantive. I discuss related issues, including the distinctions between pluralism, deflationism, and racial scepticism, and the presuppositions of the standard debate regarding natural kinds.

Impact Statement

This thesis argues that some philosophers are focused on the wrong questions to do with race. If correct, and if accepted, this would have the implication that those philosophers of race should turn to consider other questions concerning race. In particular, the focus would be on normative questions concerning how we ought to think and talk about race, and on the descriptive analysis of ordinary race talk and thought. Progress on these issues within academic philosophy might allow philosophers to usefully contribute to urgent public conversations around race and racism.

Another strand of my argument is that the questions which some philosophers tend to focus on are to be settled by those working in other disciplines. In particular, the questions seem to be questions for biologists, social scientists, sociologists, and others working on the phenomena of race. I suggest that the correct answers to the question of what explains race may have important implications for the normative questions which are, in part, for philosophers, concerning what we ought to do with race talk and thought. If the conclusions of this thesis were accepted, therefore, this kind of inter-disciplinary engagement might become more common.

Table of Contents

Introduction

Our world is structured by race. Systematic differences between people according to race abound. To take an example, the income of the median black household in America is less than 60% of the median white household, and the wealth of the median black household is around 10% of the median white household (*The Economist* 2020). There is much of human life which an individual lacking such concepts as *race*, *racism*, *black*, *white*, and *Asian* would be unable to understand. They would be ill-equipped to navigate contemporary society.

Sitting beside and influencing the ordinary use of racial concepts is their use by academics working in anthropology (Wagner et al. 2017), biomedicine (Burchard et al. 2003), economics (Johnston and Lordan 2015), epidemiology (Cooper and David 1986), evolutionary history (Cavalli-Sforza and Feldman 2003), history (Omi and Winant 2014), political science (Hacker 1992), sociology (Hall 1996), and, of course, philosophy. Previously, many uses of racial concepts in these fields were what Appiah (1996 79) calls *racialist*: they held that members of different races had normatively significant differences, which were to be explained by membership in that race, which was taken to consist in the possession of a *racial essence*, a simple and intrinsic property which necessarily all and only members of that race possessed. Moreover, these uses were often *racist*: they posited or presupposed a normative hierarchy of races. Many uses of racial concepts by those outside of academia have been and continue to be racialist and racist, in part due to the influence of these academic discussions.

In recent decades, explicitly racist uses of racial concepts in academia have ceased to be predominant. While racism in academia persists (Garrod 2006), many of those involved in studying racial phenomena are interested not in providing stories which can be used to justify or obscure racism, but in bringing to light how race and racism penetrate contemporary societies. Philosophers of race have gone from endorsing and attempting to justify racism to providing increasingly critical theories of race. Since the 1990s, a debate among philosophers within the analytic tradition has emerged concerning the metaphysics of race. Philosophers such as Appiah (1996), Zack (1993, 2002), and Mills (1998) applied the methodology and

insights of analytic metaphysics and philosophy of language and science to race. The debate focused on questions concerning the existence and nature of race: do races exist? If so, what are races? What is it that makes one a member of a race?

Three broad positions developed in this debate. The first, *racial scepticism*, says that there is no such thing as race. There are no races. No one belongs to any race. Racial sceptics include Appiah, Zack, and Glasgow (2009, 2019). A core motivation for racial scepticism is the thought that racial thinking is essentially racialist. It presupposes (false) racialist claims. Racial sceptics often advocate (*racial*) *eliminativism* on this basis, the view that we should stop using racial terms and concepts.

The second broad position is (*racial* or *social*) *constructionism*. Defenders of this view claim that races do exist. Races are quite different in kind, however, from what racialism holds them to be. They are produced, in some sense, by human practices. One of the practices many constructionists point to in the production of races is the representation of people as having racial essences, or as having a race, more generally. Constructionists include Mills (1998), Root (2000), and Haslanger (2000, 2019). Many constructionists defend (*racial*) *conservationism* - the view that race terms should be retained. They often hold that they should be used differently, in the knowledge that their referents are *not* biological kinds.

The third broad position is biological realism. On this view, races exist, and they are groups distinguished by the possession of some biological property. To be a member of a race is to possess some biological property. Like constructionists, biological realists tend to be conservationists.

My view is that there is something wrong with this debate. The problem, I argue, is that the participants to it do not actually disagree about the underlying, non-linguistic facts. The disagreement concerns what *race* and other racial concepts apply to. But that's not a substantive metaphysical debate. Even where philosophers *do* differ over the underlying, non-linguistic

facts, the disagreement seems like one for biologists, sociologists, and social scientists to settle. There *are* substantive questions concerning race. They concern the explanation of the phenomena associated with race. But this, too, is a task for those working in other fields than metaphysics.

The thesis begins, in Chapter 1, with an examination of the claim that race is socially constructed. It asks what we should understand the constructionist *metaphysician* to mean by this claim. Chapter 2 lays out a proposal inspired by Barnes's recent work on the metaphysics of gender (2020), and disability (2016). Barnes advocates two core theses, the *negative thesis* and the *positive thesis*. The negative thesis as applied to race says that the metaphysics of race need not line up with the extensions of ordinary language racial terms or with ordinary race concepts, nor with true ordinary language sentences or ordinary racial beliefs. Metaphysics need not be in the business of conceptual analysis or of spelling out what it is to satisfy a natural language term. I argue that realists about metaphysics should accept the negative thesis. The positive thesis concerns what a metaphysics of race should achieve, if not those things. Barnes's positive thesis as applied to race is that the metaphysics of race should say what explains race. I argue that there is no distinctive explanatory task for the metaphysician of race. The task of explaining race is for other disciplines. The correct metaphysics of race will be decided by which of these disciplines is able to explain the phenomena of race.

Chapter 3 advocates *deflationism* with respect to the metaphysics of race debate. I argue that there is limited disagreement between the participants of the actual metaphysics of race debate about the underlying, non-linguistic facts. Where there is such disagreement, it is not up to metaphysicians of race to settle it. I then argue that there are important and interesting questions concerning racial thought and language, both descriptive and normative. The normative question concerning what we ought to do with racial talk does not, as is often taken to be the case, directly depend on the metaphysics and semantics of race.

Chapter 1: "Race is Socially Constructed"

Thinkers have advanced myriad theses explicitly couched in the language of social construction (Hacking 1999 1). Some, such as Foucault (1972, 1978), avoid the language of construction but are nevertheless recognisably constructionist theses (Mallon 2007a 94). It is common to claim that race is socially constructed. One of the major positions in the metaphysics of race debate goes under the label 'social constructionist'. But what the claim should be taken means in the context of the metaphysics of race debate is not immediately clear. Is the thought that there is a radical dependence of world on mind? Is it that races are the causal products of social factors? Or is it that race is somehow a product of the social but is nevertheless objective and mind-independent, in some relevant sense?

The aim of this chapter is to locate the best interpretation of the *social constructionist claim* - the claim that race is socially constructed. The best interpretation of this claim is the one with the best prospects of vindicating the metaphysics of race debate. By 'the metaphysics of race debate', I mean the debate inaugurated and participated in by Appiah (1996), Mills (1998), and Zack (1993). To *vindicate* this debate would be to give a specification of what the social constructionist claim means meeting at least the following conditions:

(1) The social constructionist claim so interpreted should be a thesis with which the other parties to the debate disagree, or with which they should disagree.
(2) This disagreement should be genuine and substantive, not merely verbal.
(3) The social constructionist claim so interpreted should be (fairly) closely related to theses that are generally labelled "social constructionist", or that are paradigmatic social constructionist theses.

What justifies these conditions? (1) ensures there actually is disagreement between the parties. (2) ensures that this disagreement is not merely terminological. (3) is justified because an interpretation on which the social constructionist claim isn't so related to work that is generally

considered to be social constructionist runs the risk of failing to be a genuine form of social constructionism at all. Note that the interpretation need not be so related to *all* projects which go under the label 'social constructionist'. Given the variety of such projects, this would be unduly, perhaps impossibly, demanding.

I will argue that a *constitutive* interpretation of the social constructionist claim has the best prospects of meeting these three conditions. It makes sense to interpret the participants to the metaphysics of race debate as primarily arguing about the conditions on being a race and on being a member of a race. Chapter 3 will argue that there is nevertheless no genuine, substantive disagreement about the underlying, non-linguistic facts. But it is fair to argue this only after trying to find the claim which has the best prospects of vindicating the debate.

1.1. The Idea of Social Constructionism

If there is a core idea associated with social constructionism as applied to people, it is that some of the concepts by which people are labelled, and some of the categories to which we belong, are the result of human decisions, practices, and institutions (Mallon 2016 4). They aren't inevitable. Our concepts do not, or did not when introduced, reflect divisions "out there" in nature. There were no such divisions to be discovered, existing independently of humans' coming to represent them. For most metaphysicians, there is a disanalogy to be drawn between divisions like this and, for example, the division of people into those with and without a certain gene. Those are divisions which are out there for humans to discover before we draw them. The target interpretation of the social constructionist claim, then, should fit with this notion, in order to satisfy condition (3).

The idea that something is the product of social practices and human decisions could equally apply to *obvious* social artefacts such as sports teams or bodies of law. They are *obvious* social artefacts (for us) because, on reflection, it is clear that they are the products of social practices and human decisions. But constructionist theses are typically advanced with respect to

categories of which a further condition is taken to obtain. This is the condition that it isn't already widely recognised that these categories are the product of human activities. Unlike a sports team or a legal corpus, many (still) take race to be natural and, specifically, biological (Shulman and Glasgow 2010 253).

As Mallon (2019) suggests, it is helpful to think of social construction as a two-place relation: X constructs Y. A precise interpretation of the social constructionist claim should specify what does the construction (the X-value), what is being constructed (the Y-value), and what the construction consists in. The claim that race is constructed could be the claim that our *representation*s are constructed, or it could be the claim that the worldly categories of which *race* is a concept are constructed. Our *representations* include, according to Mallon (2016 6), the racial concepts, beliefs, ideas, theories, attitudes, narratives, models, pictures, norms, rules, utterances, inscriptions, and texts by which we represent the world and which structure our actions. By 'structure our actions' Mallon intends Anscombe's idea of acting *under a description* (Anscombe 1957 11). The racial categories themselves, by contrast, are the things which these representations are representations *of*, or which they are *about*. As I lay out more fully below, the distinction between the construction of races and of individual racial identities - *being* black, white, and so on - is not always sufficiently noted. Bearing this distinction in mind will reveal interesting results about what the different parties to the debate are committed to.

The discussion in the remainder of this chapter is structured by considering different candidate objects of construction, within which different candidates for the construction relation are considered. I argue that the pattern of compatibilities and incompatibilities shown in the table below holds. The best explanation for this is that a constitutive interpretation of the claim is correct, since if so it will entail the constructionist's commitment to the social causation of individual racial identities and will either be or imply or be roughly in the vicinity of the four last four claims in the table. My discussion of what *does* the construction and, if it is done with a purpose, to what purpose, is more limited. This is not because it is a less interesting question, but because it seems to be a question for historians, sociologists, and social scientists. Chapter

2 will argue that *all* these questions are to be settled by research in these other fields, rather than by metaphysicians. But first we should examine what questions the metaphysicians actually set about trying to answer, and what their answers are.

Claim	Constructionism	Biological Realism	Racial Scepticism
Our racial concepts' (application conditions) depend on our using them (in certain ways/for certain purposes) (1.2.1)	Compatible	Compatible	Compatible
There is no mind-independent, objective reality concerning race (1.2.2)	Compatible	Compatible	Compatible
Idea-constructionism about *race* (1.2.3)	Compatible	Compatible	Compatible
Racial identities are socially caused (1.4.1)	Committed	Incompatible	Incompatible
Racial identities are biologically caused (1.4.1)	Incompatible	Committed	Incompatible
Races are socially caused (1.4.2)	Compatible	Compatible	Incompatible
Races are Searlean social institutions (1.5.1)	Implausible	Incompatible	Incompatible
Races are Lewisian conventions	Implausible	Incompatible	Incompatible

(1.5.1)			
Races are objective kinds which would have no independent epistemic interest were it not for our practice of classification (1.5.2)	Plausible	Incompatible	Incompatible
Races are kinds produced by the causal effects of our systems of classification (1.5.2)	Plausible	Incompatible	Incompatible
Races are explanatorily significant kinds partially constituted by our practices of representation (1.5.2)	Plausible	Incompatible	Incompatible
Races are kinds which are widely taken to be natural but are in fact social (1.5.2)	Plausible	Incompatible	Incompatible

1.2. Our Representations are Socially Constructed

In this section, I explore interpretations of the social constructionist claim which are about our representations of race. The idea is that human actions, practices, decisions, and institutions construct our representations of race, including beliefs about race and racial concepts. A few different versions of this claim are discussed. Hacking's *idea constructionism* constitutes a sophisticated version of this interpretation. While Hacking developed this notion to apply to debates about the construction of scientific theories and entities - the so-called "science wars" - it could be applied to the construction of human category concepts, too.

However, I'll argue that neither this nor any other thesis primarily focused on our racial representations is what is at stake in the metaphysics of race debate. None of these is an interpretation that will vindicate the debate.

1.2.1. Counterfactual Dependence

One version of this claim is a counterfactual dependence claim: if humans didn't talk and think, then we would not have concepts like *race*, *black*, *white*, and *Asian*, and we would not have racial beliefs. This claim is uncontroversial. It doesn't distinguish *race* from any other concept. Concepts like *knowledge*, *electron*, and *dog* are all such that if we didn't think using them and if we had no words which expressed them, then we would not possess them. If we possessed no racial concepts, then we would not be able to form beliefs which essentially involved them. This cannot be an interpretation of the claim which will vindicate the debate, then. It is uncontroversial (violating constraint (1)) and insufficiently related to paradigmatic social constructionist work (violating constraint (3)).

Perhaps by specifying particular ways of thinking, talking, and behaving that our concepts and their application conditions depend on can create a more controversial claim, more in-line with paradigms of social constructionist work. Or perhaps the same could be achieved by specifying particular purposes for which these concepts were introduced or given the application conditions they have. Mills, for example, makes the claim that what American racial concepts apply to was determined to "establish and maintain the privileges of different groups" (1998 48). And this is a general idea familiar from Foucault (1972), although in Foucault's thought the role of individual agency is insignificant. This would certainly yield a stronger claim. Not all concepts' application conditions are or were determined in the service of certain groups' interests.

But while it is controversial, it isn't the right controversy to vindicate the metaphysics of race debate. The debate is not over *which* social factors played a role in how we divide people into

races, or to what end. Biological realists can agree with Mills that the application conditions for racial concepts *were* determined by the privileged, or in their interests. They can hold that they are nevertheless biological categories. Perhaps there are biological kinds which it is in the interests of the privileged to recognise, and the recognition of which served to ideologically undergird a racist system which privileged members of one kind and subordinated others. Racial sceptics can also agree with this. Perhaps social factors contributed to the development of the race concept's application conditions, which are biological but not actually fulfilled by anything. Indeed, this claim is part of Appiah's (1996 100-101) argument. So, this interpretation doesn't vindicate the metaphysics of race debate.

What about other representations, in particular our beliefs and theories about race? Could the debate be about what caused us to adopt certain beliefs and theories (either at all or rather than others)? This interpretation is more viable, and will be explored as part of the discussion of Hacking's idea constructionism.

1.2.2. Global Constructionism

What we have terms and concepts for is determined partly by our interests. How we divide the world is partly up to us, even if, as realists like Sider (2009) hold, there are better or worse ways of doing so. Social factors influence the contours of our concepts. Some philosophers, such as MacKinnon (1989 114), apparently hold that because our representations of the world are governed by social factors, we don't have access to any mind-independent reality (in a particular domain). Therefore, reality itself (or that domain) is constructed.

This kind of anti-realist interpretation is not what is at issue in the metaphysics of race debate. While the claim is controversial, it is orthogonal to the debate. Social constructionists can be realists. They can accept that we know a lot about mind-independent reality, and that there are mind-independent facts about race, in the relevant sense. For example, on Haslanger's (2000) theory of race, one's being perceived to have certain features indicative of certain continental

ancestry which plays a role in one's systematic privileging or subordination is what it is for one to be a member of a given race. While one's race is mind-dependent in the sense that it depends on how one is perceived, it is an objective fact that one is a member of a given race, and a belief about it is objectively true or false, depending on what it says and how the world is. Conversely, a biological realist could be an anti-realist by believing that race is a biological property, but that the biological world is constructed. Radical or global constructionism is not what is at issue.

1.2.3. Hacking's Idea Constructionism

Hacking (1999) identifies an interpretation of the claim that a representation is socially constructed which gets closer to the particular purpose of much constructionist discourse. Constructionist work, for Hacking, is critical: it stresses the contingency of current arrangements, claiming that there are some bad effects of these arrangements, and that we can and should change our practices, which would consequently change how things are. For such a thesis to be purposeful, it must be that the contingency of current arrangements on human practices and social factors isn't widely recognised. So, there couldn't be a (purposeful) constructionist thesis regarding obvious social artefacts, like sports teams, since it is a presupposition of such a thesis's being purposeful that the object of construction's being socially produced not be widely acknowledged. As Haslanger (2012a 116) points out, concepts in general are obvious social artefacts. Indeed, representations more broadly are obvious social artefacts, given that they depend on the existence of socially produced conceptual schemes. We know that concepts are often culturally specific and could easily have been otherwise. They are not foisted onto us by the universe. So what room is there for a constructionist thesis concerning a representation if purposeful constructionist theses do not concern obvious social artefacts?

Hacking identifies a cluster of three theses concerning a particular kind of representation which make for a purposeful constructionist thesis. This particular kind of representation is

an *idea* or *theory* (Hacking 1999 22). An idea is (roughly) a conception - a concept with associated beliefs. The *idea constructionist* makes three contentious claims with respect to an idea. First, they claim that our understanding of the domain, D, which the idea is about is contingent (1999 68). Second, they endorse *nominalism* about kinds in D (1999 80). Third, they explain the stability of our understanding of D in external rather than internal terms (1999 84). Hacking's notion is that a constructionist thinks that how we divide up a domain and the ensuing content of our beliefs is contingent; that there are no natural kinds which our classifications (should aim to) latch onto; and that the stability of our classificatory schemes and beliefs is to be explained, for example, sociologically, rather than by reference to the stability of the domain itself.

Consider the example of physics. The contingency claim denies the following: any successful physics which explains our world would have to develop in much the same way as ours did. It would have to be *equivalent*, in some sense, to our own. If the contingency thesis is true, then an alternative, equally successful physics is possible. It is difficult to pin down precisely what 'equivalent' or 'successful' mean here. Hacking employs the notion of a *robust fit*, drawn from Duhem (1906/1954). This is where theory, apparatus, theoretical understanding of the apparatus, and interpretation and analysis of experimental results come together to produce a stable equilibrium. The theory succeeds because it is stable in light of experimental results obtained using the apparatus, interpreted and analysed in light of the theoretical understanding of the apparatus. The contingency thesis does not say that our physical theories aren't successful or true. It says that a genuinely distinct physical theory could have been equally successful and true. Alternative robust fits could have developed.

The nominalist thesis denies that the "facts are there, arranged as they are, no matter how we describe them" (Hacking 1999 83). The nominalist thesis says that the world "does not even have what we call structure in itself" (ibid.). Hacking illustrates this with reference to the term 'Douglas Fir'. The nominalist holds that "there is nothing peculiar to the items picked out by a common name such as "Douglas Fir," except that those items are called "Douglas Fir.""

(ibid.). Nominalists think that nature has no joints to be carved. Anti-nominalists think that there are privileged divisions picked up on by terms like 'Douglas Fir', already out there in nature.

Finally, there is the explanation of stability. The constructionist "holds that explanations for the stability of scientific belief involve, at least in part, elements that are external to the professed content of the science." (Hacking 1999 2). Typically, these are social factors. The idea is that our physical theories don't change all the time, not because they are accurate theories of a stable universe, but because of social factors such as the charisma and power of influential theorists.

Haslanger (2012a) criticizes idea constructionism as an interpretation of the social constructionist claim. While Haslanger's criticisms are, I shall argue, mistaken, considering them will allow a sharper understanding of what Hacking intends. The question, according to Hacking, is whether our concepts' origins and retention are determined by the domain of which they are concepts. The idea constructionist says that they aren't, and the non-constructionist says that they are. By 'determine' here, Haslanger takes Hacking to mean causal determination: the non-constructionist holds that there is an inherent structure to D (denying the nominalism about kinds in D) which causally determines our understanding of it, and the stability of this structure together with its causal determination of our understanding of the domain ensures the stability of our understanding of D.

Haslanger thinks Hacking's non-constructionist view is implausible (2012 118). She claims that few would hold that their concepts or classifications are *inevitable* given how the world is. Indeed, the fact of conceptual schemes' varying between cultures and individuals seemingly refutes this. One need not adopt anything so radical as nominalism about kinds in a domain, or hold that the stability of our understanding of the domain is to be explained *purely* in external terms, or hold that everything about our understanding of the domain is contingent, to hold that our concepts of a domain are not inevitable given how that domain is. As Haslanger

writes, our classifications are "the product of some combination of worldly input from perception and experience and social input from language, practices, and the like" (2012 119). Some compromise between the extreme constructionist view which Hacking elaborates and the view which Hacking has opposed to it is much more plausible than either extreme, according to Haslanger.

This criticism relies on an uncharitable reading of Hacking. The debate according to Hacking isn't between these two "extreme" positions that Haslanger identifies. The opponent of the constructionist, according to Hacking, does not have to say that our ideas are *causally determined* given the inherent structure of D. The anti-contingency claim is a more moderate one that any pair of equally successful theories would be equivalent. The idea constructionist denies this. Perhaps the misunderstanding relies on the slipperiness of the notion of equivalence. It is plausibly at least a sufficient condition on two theories' being equivalent that there be a straightforward translation of one into the other. It is presumably not a necessary condition on equivalence, however, that they have exactly the same concepts. Thus, there is room for conceptual variation. The anti-constructionist claims that the reason our understanding of domain D is stable is that D itself is stable. Our theories are (often) broadly correct, our methods broadly successful. Because the domain is stable, our theories will therefore be stable. This doesn't mean that the form and content of our theories is fully causally determined by the domain. The constructionist denies that this is the correct explanation of the stability of our understanding of domain D. They hold that what keeps our understanding of D stable is something else, something social. On this reading, Hacking's non-constructionist view is much more plausible. The constructionist view is also readily defensible while being radical, which is to be expected. Haslanger's criticism, then, is mistaken.

Even with this more charitable reading of Hacking's idea constructionism, interpreting the social constructionist claim as this cluster of theses about the idea(s) of race does not vindicate the metaphysics of race debate. An initial problem is that the notion of success does not obviously carry over to non-scientific "theories". It is unclear what a successful folk theory of

race would be. The idea of a non-equivalent but equally successful folk theory is even more obscure. This is not a criticism of Hacking, who developed idea constructionism to understand debates over the construction of scientific theories.

Putting these problems aside, this interpretation fails to satisfy constraint (1) because all the sides to the debate can agree that Hacking's non-constructionist's view is false. The biological realist can agree that our ideas about race are not explained by the stability of the structure of the racial domain, particularly our ordinary, non-scientific, ideas. A biological realist such as Andreasen (2000) can allow that ordinary ideas about race are the products of social factors. Andreasen argues that biological realism is compatible with the view that "biology lends no support to [ordinary] beliefs about ... race" and that we can "explain these false beliefs by appeal to ideological factors" (2000 662). She can agree that ordinary concepts of race and the beliefs formed using them persist because of social factors, rather than the domain which they purport to be about. People are brought up to believe in races, and this explains our retention of race concepts and beliefs. The biological realist can say that there are nevertheless biological kinds which biologists find useful in explaining various phenomena, and that these count as races. If biological realists can accept these two points, it leaves only nominalism about the racial domain as a possible division between them and the constructionist. But this issue is orthogonal. As will be seen later in discussing Haslanger's theory, social constructionists can endorse the existence of robust racial kinds, holding that members of a race share something in common beyond being labelled as members of the same race. And a nominalist can be a biological realist. Nominalists deny the existence of inherent structure. To be a biological realist, one can hold that there are many ways of dividing the domain for biology, even though these biologically interesting divisions don't track metaphysically privileged structures (Kitcher 2007).

Perhaps this has moved too quickly. The biological realist will surely want to maintain that in order to understand a certain domain - the domain which the biologist supposedly needs the (biological) concept of race to explain - one will have to adopt a certain theory of race. Any

other successful theory of this domain will be equivalent to this theory. That is, any successful theory will make reference to races. And presumably the biological realist will claim that stability of understanding of this domain is explained by the stability of the structure of that domain. If the constructionist denies any of *this,* then there appears to be a substantive debate in the offing. The short response to this is that social constructionist metaphysicians simply do *not*, in general, deny that there are some domains the understanding of which requires a theory utilising concepts equivalent to the race concept as understood by the biological realist. The debate is just not about nominalism concerning certain biological domains, or about how biological theories' stability is explained. Constructionist metaphysicians can agree that the entities and kinds which the biological realists hold to be the referents of (biological) race concepts exist. They can accept that the stability, such that there is, in biologists' understanding of these categories is explained by the stability of the categories themselves. They just deny that these categories are races. This will be more fully explained in the discussion of deflationism with respect to the metaphysics of race debate and the ontological consensus (Mallon 2006) in Chapter 3.

The biological realist can accept the three theses which make up idea constructionism concerning race. Can the racial sceptic? They can agree that our understandings of race are contingent, because they don't even grant that they are successful. They will likely, as Appiah (1996 104) does, appeal to social factors to explain the widespread adoption of what is in their view mistaken beliefs about human categories. It is less clear what to say as regards nominalism concerning the domain of race. For a typical argument for racial scepticism is that racial talk is committed to the existence of certain structures, biological kinds with shared essences linked to normatively important characteristics, and that the non-existence of these structures undermines racial talk. But a nominalist won't generally be an error theorist. They will instead deny that ordinary talk does indeed presuppose the existence of such structures. Nominalists need not hold that 'There are Douglas Firs' is false. They need only deny that there is anything that Douglas Firs have in common, beyond being labelled "Douglas Firs". In that case it would look difficult for the racial sceptic to accept nominalism. But there are other ways of being a

sceptic which are compatible with nominalism. For example, one could hold that whereas biologically useful categories support generalizations and explanations, racial categories do not. Or, as Zack does, one could argue that racial concepts have inconsistent application conditions, and are therefore "unfounded and unfair" (1993 5). The presuppositions typically made in the debate concerning natural kinds will be discussed further in Chapter 3.

I've argued that Hacking's idea constructionism won't vindicate the metaphysics of race debate. It fails to identify a controversy between different parties to the debate. It is unlikely that some other interpretation of the social constructionist claim which takes it as a claim about the construction of our *representations* will vindicate the metaphysics of race debate. This is because metaphysics generally is not primarily concerned with concepts and representations. It is concerned with the domain that the concepts are about, or purport to be about. Metaphysicians of race, unlike many constructionists working in other fields, are not concerned with the genealogy of our ideas about race, except insofar as this might tell us something about races themselves. The focus for metaphysicians is the world, not our representations of it.

1.3. Construction of Races and Racial Identities

Metaphysicians are interested, not in our racial representations, then, but in races themselves, and in individuals' racial identities. I use the term 'racial identities' or 'membership of a race' to avoid using the potentially confusing term 'an individual's race' to refer to someone's being white, black, and so on. Consider the category of white people. To say that this category is socially constructed is to ascribe to it a certain relation with human decisions, institutions, actions, and practices. To say that an individual's racial identity, or membership of a race, is constructed is to say that that individual's being white or black or Asian is related in some way to certain human decisions, institutions, actions, or practices.

The remainder of this chapter explores different accounts of the construction relation for races and for individuals' racial identities. Section 1.4 deals with causal interpretations. I argue that different parties to the metaphysics of race debate *do* have certain commitments regarding the causal origins of racial identities. This is explained by their holding different accounts of what constitutes race and racial identity. Section 1.5 discusses constitutive interpretations. Section 1.5.1 deals with possible theories on which constructions are social institutions, in Searle's (1995) sense, or conventions, in Lewis's (1969) sense. I'll argue that neither will work. Section 1.5.2` gives four different kinds of constitutive constructionist accounts of race. I argue that if there is an interpretation which vindicates the metaphysics of race debate, it is a constitutive one.

1.4. Causal Construction

Causal accounts of the construction relation say that human decisions, practices, actions, and institutions cause races or individual racial identities to exist, persist, or have the features they have. This section explores such accounts, arguing that interpreting the social constructionist claim as a causal claim fails to fully explain the disagreement between the different parties to the metaphysics of race debate.

1.4.1. Causal Construction of Individual Racial Identities

One promising avenue for constructionist analysis is how individuals come to have their racial identities. Haslanger discusses *discursive construction*: "something is discursively constructed just in case it is the way it is, to some substantial extent, because of what is attributed to it" (Haslanger 2012 88). All of us are, to a significant extent, how we are because of what is attributed to us. Interpreting this 'because' causally, the claim is that having certain properties attributed to us causes us to develop certain features. For example, children who are labelled "academic" may get more attention from teachers, they may be encouraged to read more, and they may become more confident in their academic abilities and come to value intellectual

achievement. Being labelled "academic" affects how they think of themselves and how others think of them, thereby affecting how they behave and how others behave towards them. Children whom other children see as "weird" (perhaps for some non-social feature such as appearance) may be socially ostracised, leading to their failing to develop the usual social abilities of children their age. Thus, they may come to increasingly fit the label "weird".

Applying this to race, the thought is that being labelled "white", for example, causes one to develop certain features the possession of which is sufficient to be white. The labelling itself *causes* the development of the racial identity. Perhaps being labelled "white" leads to one being thought of, by oneself and by others, in certain ways. This affects how one is treated and how one treats others. This has social, political, economic, and cultural effects. Norms, expectations, and constraints are placed on one, communicated and enforced by communities with which one interacts.

It seems plausible that being labelled "white", "black", or "Asian" can significantly affect one's life, whether or not it causes one to *be* white. As an interpretation of the social constructionist claim, though, the claim must be that it causes one to *be* white, just as being labelled "weird" may cause a child to *be* weird. This is because the claim that being labelled "white" significantly affects one's life is consistent with racial scepticism and with biological theories of race. It is consistent with racial scepticism because we may attach normative significance to features which we mistakenly take people to have. It is consistent with biological theories of race because being given a biological label (accurately or not) may have normative significance attached to it by society. Having a certain eye colour, for example, could have normative significance attached to it (as in Jane Elliot's famous "Blue eyes/Brown eyes" exercise), but this would not suggest that eye colour is socially constructed.

The claim that being labelled "white" *causes* one to *be* white, by contrast, is inconsistent with racial scepticism and biological realism. Biological realism holds that membership of a race is like being a member of a certain subspecies, a biological feature. But it's not plausible that an

animal's subspecies membership is caused by being labelled in certain ways. So, membership in a race cannot be caused by being labelled in certain ways.

Kitcher (1999) allows that segregation may have produced reproductively isolated populations to which 'race' applies. Could this mean that labelling could in some way cause an individual to have the racial identity that they have, by causing such segregation? No. The story isn't that one is born raceless, and that subsequent labelling using one racial term rather than another causes one to develop features such that one has a certain racial identity. The story is that labelling produced a reproductively and geographically isolated *group* - a race. The individual's racial identity is determined biologically. While biological developments might have causes ultimately rooted in human practices of labelling and the attachment of normative import to the labels, these practices don't directly cause an individual to be a member of the particular race they are a member of in the way that the social constructionist suggests. Consider the example of dog breeds. These have social causes, being rooted in human decisions and interests. But a particular dog's breed isn't socially caused. That they are a member of one breed rather than another is determined biologically.

If one is caused to be white, then one is white, and racial scepticism denies that anyone is white. So, any claim that people are caused to be members of races is straightforwardly inconsistent with racial scepticism. One (odd) way to be a racial sceptic would be to say that people *would* have racial identities if certain social facts obtained which actually do not. But this is not a strategy I have seen pursued, and any plausible candidate social factor for causing race surely does actually exist. Racial sceptics normally hold that 'race' would refer to a biological property, if it existed, but, because no such biological property exists, it does not. In that case, racial identities couldn't be caused by social factors, for the same reasons as the biological realist's. So, the constructionist claim interpreted in this causal way is inconsistent with racial scepticism.

The term 'labelling' has been used. This suggests there must be concepts with which to label people for there to be this kind of causal effect. Haslanger appears to endorse this thought:

> discursive construction depends on there being descriptions, distinctions, and classifications at hand whose attribution to things makes a difference - I am the way I am today because people have had the linguistic and conceptual resources to describe me as, for example, "smart" or "stupid," "attractive" or "ugly." (2012 89).

This is familiar from Hacking (1995) who relies on Anscombe's (1957) influential view that actions require that the agent be acting "under a description". But it is a mistake to place too much emphasis on the availability of concepts, as Haslanger (2012a) acknowledges in a different essay:

> there may not be an explicit or named social category … yet in developing a social theory for the society in question, it may be important to introduce such a category in order to understand ... the outcomes one is concerned to explain. (2012 136)

Haslanger, to be fair, does not require that discursive construction involve only spoken language in her discussion of it (2012 88 n.15). But there may not even be a publicly available *concept* which is used to label people, while much the same process goes on. Children could ostracize and thereby significantly affect a different-looking child without having a concept or word for either their physical difference or their imposed social status as an outcast. It is important not to intellectualise the process of construction without due justification here.

Such a justification might be thought to be provided by a special phenomenon which Hacking (1995, 1999) identifies that occurs where a label *is* available. This occurs where people falling under a concept react to falling under that concept, thereby altering the meaning of the concept. Hacking calls this a "looping effect". This explains the relevance of Anscombe's account of intentional action, and provides one way of understanding otherwise extremely

25

puzzling Foucauldian claims concerning the emergence of homosexuality in the nineteenth century: only when the term 'homosexual' was introduced to medicalise certain behaviour did there emerge a kind of person with certain distinctive features corresponding to the term, in part because only then did people begin to think of themselves and their actions in certain ways.

This still shouldn't prevent us from acknowledging that people can be affected in the relevant way by behaviour which is coordinated *without* a label. What is important to the kind of causal process under discussion is that a group coordinate its behaviour in treating one another in certain ways, whether that is sanctioned by a label imbued with normative import or not. Indeed, even looping effects may occur with or without a label. Hacking discusses autistic children reacting to being categorised as autistic without knowing that they are labelled using the label 'autistic'. In this instance, humans are reacting to being labelled without knowledge of the label or of their being labelled. It seems possible that there may be coordinated behaviour without *anyone* having a concept to label the individual or group of individuals in question. A child who is ostracised because of their physical appearance may be significantly affected by that. Such children may form a kind, supporting explanatorily significant generalizations. Aside from the metaphysics, it is important to acknowledge this as a possibility, for there are many cases of oppression where there is no explicit labelling, which makes the oppression harder for the oppressed to resist, a situation Fricker (2007) labels *hermeneutic injustice*.

The claim that being labelled, or otherwise differentially treated, as a member of a given race causes one to be a member of that race is not a claim which the biological realist or the racial sceptic can accept. But this interpretation still fails to fully capture what is at issue between the parties to the debate. It appears to meet constraint (1), in that the parties to the debate apparently disagree over it. We have seen no reason to think the debate is non-substantive, so have no reason yet for thinking that it fails to meet constraint (2). And it would appear to be in line with paradigmatic social constructionist theses, meeting constraint (3). There is a

significant strand of recognisably constructionist theorizing concerning how individuals acquire their properties. Examples include Foucauldian projects such as Hayward's (2000) study of how teachers at two different schools with children coming primarily from working class areas in one and middle class areas in another enforce different norms that significantly affect how the children conceive of themselves and others, and thereby affects who they are and how they behave. But while this interpretation meets these constraints, it does not exhaust the disagreement. In fact, the disagreement can be seen as deriving from a logically prior one. The biological realist holds that racial identities can't be caused by social factors because they are biological properties. The constructionist holds that racial identities can be so caused because they are social properties. The sceptic denies that they are caused by either because there are no such (instantiated) properties. These are different views about what it is to have a certain racial identity.

1.4.2. Causal Construction of Races

A distinct thesis is that the existence, persistence, or character of races is caused by human actions, decisions, institutions, and practices. Some biological realists, such as Andreasen, take it that races are produced by natural - meaning non-social - factors. Andreasen (2000, 2004, 2005) identifies races as *clades*, reproductively isolated human breeding populations which are *monophyletic groups*: they are made up of an ancestor and all of their descendants. It is implausible that such clades have developed recently as a result of human decisions and practices. They would have developed early in human evolutionary history. The differential evolutionary pressures on such isolated populations presumably explain the correlation of morphological features and other phenotypic differences with continental ancestral origin. But, as noted earlier in the discussion of Kitcher (1999), other biological theories of race can allow that they *are* socially caused. That Kitcher doesn't require that groups be monophyletic means that no significant evolutionary time needs to have passed for them to have developed in, in contrast to Andreasen's (2005) theory. So, some but not all biological realists hold that races are produced by natural processes, although even Andreasen's clades could have been produced

or maintained, or could in the future be produced, by human decisions, given sufficient time. In short, the biological realist is not committed to the non-social origins of races.

Social constructionists *appear* to be committed to the social origins of races. They need not deny any of the evolutionary history which Andreasen takes to provide a basis for racial classification. They can accept that differential evolutionary pressures on relatively reproductively isolated groups produced geographical clusters of phenotypic traits. They should seemingly deny that such groups are races, and that races were caused to exist by natural rather than social processes. But these appearances are misleading. Just as some biological realists can accept that groups characterized by biological properties were caused to exist by human decisions, actions, and practices, some constructionists can accept that races were caused to exist by natural processes, such as moderate reproductive isolation. To leap ahead somewhat to constitutive accounts of the construction relation, consider a simplified version of Mills's (1998) account. On this version, what it is to be black is to have certain morphological features. This is because these are the application conditions we associate with the concept *black*. What it is to be white is to have certain other morphological features, those we associate with the concept *white*. What makes this constructionist, as will be discussed later, is that these kinds are, or would be, of no independent explanatory interest, absent this conceptualisation. One could say that these groups were caused to exist by differential evolutionary pressures. If this is right, then even some forms of social constructionism - "thin" ones - can accept that races were produced by biological or natural processes.

So biological realists aren't committed to the non-social origins of races. And some constructionists can accept the non-social origins of races. Most biological realists *will* hold that races are the products of biological processes. Most constructionists will hold that races are the causal products of social processes. But, as far as metaphysicians are concerned, there is no commitment to these claims. Racial sceptics, of course, deny that anything causes races to exist. But, nevertheless, the causal construction of races can't be what is at issue in the metaphysics of race debate.

1.5. Constitutive Construction

In the last section, I argued that versions of social constructionism and biological realism can accept biological causes of races and social causes of races, respectively. The metaphysics of race debate is not about the causal origins of races. I argued that there *are* differing commitments over what causes individual racial identities. This difference of view stems from differences of view over what the racial identities themselves consist in. Proposals concerning what it is to be a race, and what it is to have a racial identity, are the subject of this section.

The idea is that certain human activities, decisions, practices, and institutions *constitute* race and racial identities. These social factors somehow produce or generate races and racial identities. This is supposed to be a non-causal relationship. At the very least, this means that these social factors are *necessary* for races and racial identities to exist. Mallon (2019) draws an instructive comparison: something could have all of the intrinsic features of a watch without representations expressing the concept *watch* and without any human activity regarding it, or watches generally. A chance physical happening could just result in an object with all of the intrinsic features of a watch. Cocktail parties and wars, on the other hand, are such that without certain representations expressing the concepts *cocktail party* or *war* they couldn't have been *cocktail parties* or *wars*. Or so, at least, Searle (1995 33-34) and others have claimed. Foucault (1978) claims that the same is true of homosexuality, and concludes that there were no homosexuals before the concept *homosexual* was expressed in the nineteenth century. The constructionist thinks that a similar relationship holds between our representations (and social practices more broadly) and race and racial identities. As Mallon suggests, what is wanted "is, at a minimum, a model of this production" (2019). The rest of this section is an exploration of various such models.

1.5.1. Searlean Institutions and Lewisian Conventions

There are already well-developed accounts of (parts of) social reality. One is Searle's (1995) account of social facts and institutions. Another is Lewis's (1969) account of conventions. Given this, and given constructionists' habit of utilising explicit comparisons between obvious social artefacts and non-social entities to make their theses more intuitive (for example, Haslanger 2012 185), it is tempting to simply apply one or both of these accounts to covert constructions. *If* race is constructed, then it is covertly constructed: it is not widely acknowledged or recognised to be a product of human decision, practice, or institution. It is widely believed to be natural. So, if these accounts cannot be applied to covert constructions, as I will argue, the prospects of using them to explain race are poor.

On Searle's account of social facts, they arise out of collective intentionality. Some phenomena have a function collectively imposed on them. This occurs when there is a collective acceptance of a statement of the form:

"X counts as Y in C"

When there is such collective acceptance, this is a *constitutive rule* for being an instance of Y. The term "X" specifies the phenomena on which the *status function* is imposed. A status function is a function something can fulfil (only) by being granted a certain status. The status function is not such that something satisfying "X" would *already*, without its collective imposition, be able to perform it. Collective acceptance is an instance of *collective intentionality*. This is intentionality by a group (a "we-intention"), not reducible to any (aggregate of) individuals' intentions ("I-intentions"). The imposed status function has "deontic powers", meaning it carries and implies social obligations.

Consider money. Searle's account of money and of the fact of some particular object's being money is that we collectively accept certain conditions on being money. We collectively impose

the status function of money on things satisfying certain criteria. Focusing on physical cash, pieces of paper or metal treated in certain ways by the appropriate authorities are collectively accepted as money. The various deontic powers money has are the result of our collectively accepting that stuff satisfying these conditions has that status. An important point is that this (typically) applies to *types* rather than to *tokens*. We collectively impose this function on everything and anything satisfying these conditions. Then anything which does satisfy these conditions is money, whether or not we recognise it. Anything which doesn't is not, even if we think it satisfies these conditions or that it is money. This explains the possibility of convincing counterfeiting. Even if *everyone* believed a particular piece of paper which *didn't* satisfy the relevant conditions was money, it would still not be money on this account. Labouring the point somewhat, such pieces of paper couldn't fulfil this function without our imposing it. Nothing in the intrinsic properties of paper means that it can fulfil the function of money without our so imposing it.

This is to be contrasted with Lewis's (1969) account of convention. For Lewis, a convention is a regularity of behaviour among a population in a (type of) situation such that it is common knowledge that: (i) everyone conforms to the regularity; (ii) everyone expects everyone else to conform to it; (iii) everyone prefers conforming to it on the condition that the others do.[1] Condition (iii) says that it is a solution to a coordination problem, i.e., a Nash equilibrium: everyone is better off conforming to the regularity on condition that everyone else does. This, and some further remarks on how regularities and common knowledge of them arise, allows Lewis to explain a wide range of conventional phenomena.

There are objections to both as interpretations of the social constructionist claim, however. An immediate objection to Searle's is that since the social origins of constructed kinds aren't widely acknowledged, his account can't apply to such kinds. For if they were created and sustained by collective acceptance of a constitutive rule, we should be aware of that, by virtue of taking part in the collective acceptance of the rule. But Searle's account isn't committed to

---

[1] Lewis (1969) actually gives a slightly different, stronger condition, but this isn't relevant for our purposes.

people being aware of the social production of institutional facts. They might, for example, believe that money is intrinsically valuable (Mallon 2017 435). Searle says that social facts and institutions "seem as natural to us as stones and water and trees" (1995 4). What Searle *is* committed to is that collective acceptance ("we-intentions") of a constitutive rule is what creates and maintains institutions and facts.

This is the source of a more sophisticated objection. The problem is not that people must be aware of the social character of an institutional fact for it to obtain. It is that the propositional attitudes Searle requires are not plausibly consistent with the propositional attitudes people actually hold towards covertly constructed kinds (Machery 2014, Mallon 2017). Searle requires that, for each race r, there is a corresponding collective acceptance of a constitutive rule. There must be the following:

We intend that persons with properties p_1, p_2, … p_n count as members of race r in C

But people who believe that races are natural don't participate in such a collective intention. They have the simple propositional attitude of believing that persons with properties p_1, p_2, … p_n are members of race r. It makes little sense "to recognize facts that we believe are natural phenomena" in Searle's special sense of 'recognize', involving collective acceptance that something *counts* as something else (Machery 2014 98). A society of religious people, for example, who believe that marriage *just is* a union between a man and a woman do not plausibly participate in a collective acceptance of *counting* certain unions between men and women as marriages. They take this to be a natural (God-given) institution.

Mallon (2017) points out a possibility for Searle which Machery overlooks. Namely, the possibility, mentioned earlier, that there is a "transition from an explicitly imposed status function to a mistakenly imposed status function" (Mallon 2017 439). Thus, we require an argument establishing that such facts do not remain institutional facts in Searle's sense after everyone has forgotten their social origins. Mallon provides exactly such an argument. Once

the social origins of an institution are forgotten, and there is widespread belief that a category is natural, the cooperative "we-intentions" no longer obtain. Mallon suggests that what instead holds covert constructions together is the mistaken belief in their naturalness (2016, 2017). This makes them a fundamentally different kind of kind because the mechanism which ensures stability of the kind is different. Rather than being the disposition to cooperate and recall status functions imposed collectively, it is the belief that there are natural differences between the categories in question. If, for example, it came to be believed that money was a natural kind, that pieces of paper satisfying such and such conditions were intrinsically valuable, what would explain money's ability to fulfil its functions would no longer be the collective intention to count certain pieces of paper *as having* certain functions, but rather the widespread belief that money was intrinsically valuable. Machery and Mallon together provide convincing reason to think that covertly constructed kinds are not well modelled by Searle's account of institutional facts. If so, then, since races are covertly constructed kinds, they provide good reason for thinking that races and racial identities are not well modelled by Searle's account.

These arguments also apply to a Lewisian proposal. A Lewisian proposal would be that we treat some people as belonging to one kind and others as belonging to another because it is common knowledge that everyone does so, that everyone expects everyone else to do so, and that everyone prefers doing so on condition that everyone else does so. But people do *not* have the necessary expectations, as Mallon (2017 442) points out. In particular, what explains their attitudes and behaviour is not a preference conditional on others' preferences and actions. Rather, "when one takes something to have its properties as a result of some natural, intrinsic properties, one treats it as having those properties *independently* of how oneself and others regard it." (Mallon 2017 428). Abiding by behavioural regularities that discriminate between people with different features is not, then, to be explained as a solution to a coordination problem, but rather as a result of people's beliefs about naturally explicable kind-typical differences. This behaviour would *not* be conditional on beliefs about the dispositions or actions of others. And, similarly, the move to saying that these practices initially emerged as coordination equilibria and have since been forgotten to be so will not work, because they are

sustained by beliefs in natural differences, not conditional preferences. Even if other people did not share one's belief that, for example, men are more aggressive, one would still hold this belief and therefore possibly prefer men for some jobs.

1.5.2. Four Versions of Constitutive Construction

I've argued against interpreting the construction relation in the way that Searle (1995) models institutional facts or the way in which Lewis (1969) models as conventional facts. This section explores four different constitutive constructionist accounts, drawn from Mills (1998), Piper (1992), Root (2000), and Haslanger (2000, 2019). No effort will be made to decide between these different accounts as the interpretation of the constructionist claim. My claim is that *if* there is an interpretation which vindicates the metaphysics of race debate, then it is a constitutive account.

On Mills's (1998) account, constructionism is the view that "an objective ontological status … arises out of *intersubjectivity* … which, though it is not naturally based, is real for all that" (Mills 1998 48). Out of intersubjective agreement on who is white, black, and Asian, there arise objective facts about who is white, black, and Asian. There are links made between the possession of certain properties and an individual's racial identity. The properties Mills identifies are: bodily appearance; ancestry; self-awareness of ancestry; public awareness of ancestry; culture; experience; and self-identification. If someone satisfies all of these for a given race, for example, if someone "looks white", has European ancestry, is aware of this, is widely known to have this ancestry, engages in certain culture and not others, has certain experiences and not others, and self-identifies as white, then they are (objectively) white. The system is imperfect, since these criteria do not always go the same way. The criteria are culturally local, as is the relative weighting of different criteria.

One way of interpreting this is as a Searlean proposal. This is how Asta (2018 106) interprets it. On this interpretation, Mills thinks that collective acceptance of a statement of the form:

People with properties $p_1, p_2, \ldots, p_n$ count as black

makes it a constitutive rule for being black. Being black is then a status function, bringing with it certain constraints and other deontic powers. This doesn't seem to be justified by a close reading of Mills's (1998) text. Mills says that "people have believed that there are natural biological differences among races and that these differences run deeper than mere phenotypical traits" (1998 46). In discussing the thoughts of radicals, which he largely endorses, he says that having "the lines of demarcation … drawn here rather than there is a social decision ... that creates the (social) reality in question. So, the resultingly racialized world is in part theory-dependent, constituted by these very beliefs" (Mills 1998 47-48). As noted earlier, he suggests that the decision about where to draw the line was made to maintain the privileges of particular groups. The core claim is that it is *because* "people come to think of themselves *as* "raced", as black and white … [that] these categories, which correspond to no natural kinds, attain a social reality" (Mills 1998 48). This doesn't suggest that collective intentionality is necessary, as on a Searlean proposal. What is necessary is common *belief* that people who have certain properties are members of one race or another.

So, Mills should be interpreted as offering an account of what it is to be a member of a race on which racial membership is objective while being arbitrary from a biological perspective. Racial membership is determined by the widespread associations between race concepts and certain features, so that possession of those features is sufficient to be a member of that race. Thus, racial membership is as objective as the features associated with that racial concept. We can understand *why* racial membership has the conditions it has by investigating the history of these beliefs and the interests which their being widely shared serves. How are race concepts distinguished from other concepts? Asta is helpful on this point, explaining that it is constructionist "because intersubjective agreement determines the criteria for race membership, as opposed to something subject-independent, such as a mind-independent nature." (2018 106). There is no kind to which people would belong absent this intersubjective

agreement about what it is to belong to this category. Mallon adds that it is constructionist because it holds that features which are linked with race concepts are not "of independent epistemological interest (as they would be if they were causally important) but of interest only because a community's conceptual practice makes them so" (2006 535). These features take on explanatory significance only because we classify people in the way that we do.

Piper's (1992) suggests a different account of race. According to Piper, what black people have in common "is not a set of shared physical characteristics" but rather "the shared experience of being visually or cognitively identified as black by a white racist society, and the punitive and damaging effects of that identification" (1992 30-31). For Mallon (2006) this puts her view among "thicker" accounts than Mills's. People are not classified into races based on superficial features, with no independent epistemic interest. Rather, on thicker accounts, "one's race entails important facts about a person within a particular context" (Mallon 2006 535). One way of developing such an account is related to Hacking's (1999) idea of *interactive kinds*. The idea is that the causal effects of being classified and treated differentially as a member of one race rather than another results in a distinctive kind of person developing, to which our racial terms and concepts refer. These kinds are of independent epistemic interest, despite being produced in part by our representations, including our theories of race, which are themselves (in part or initially) mistaken.

Root's (2000) account of race is also thicker than Mills's. In contrast to interactive kind accounts, however, his claim is not that patterns of representations and differential treatment *cause* distinctive (and explanatorily important) kinds to develop. It is that patterns of representation (partially) *constitute* such distinctive kinds. They are logically or metaphysically necessary for races and racial identities to exist. Root compares race to crime: no one "would be guilty of theft had we not invented or recognized the relevant laws or drawn any property lines" (Root 2000 630). For Root, unlike Mills, races are necessarily thick kinds which feature in social scientific explanations: "we divide ourselves by race and zip code but only race says much about us … while we differ significantly in school attainment, marital status, labor force

participation, arrest, and conviction rates by race, we don't by zip code." (ibid. 631). Of course, we *do* actually differ significantly in all those things by zip code. But Root's point stands, which is that being located in one zip code rather than another is not itself explanatorily or causally significant. It is a proxy for genuinely explanatory differences such as wealth, provision of public services, and, of course, race. So, for Root, races are partially constituted by our racial representations, and they are causally and explanatorily significant kinds. For Mills, however, they could exist without being causally or explanatorily significant.

The racial generalizations which Root is interested in are local, both biomedical and social scientific. Race, he claims, does not travel:

> men who are black in New Orleans now would have been octoroons there some years ago or would be white in Brazil today. Socrates had no race in Ancient Athens, though he would be a white man in Minnesota. Where R is a race, a person is R at a site only if R is used there to divide people. (Root 2000 632)

There were no races in Athens because the Greeks did not divide people by race, or so Root claims. There were interesting racial generalizations to be made there (since racial categorizations are arbitrary from a biological point of view). While there could be objective racial divisions in the absence of explanatorily significant kinds, they would be different kinds of division. They would be equivalent to postcode divisions, rather than important human categories. So, on Root's theory, race is socially constructed in the sense that without representations of races and of individuals as raced in a context, there are no races or racial identities. With such representations, and with attendant norms which result in differential treatment, there can be causally and explanatorily significant kinds, susceptible to scientific study.

How is Root's theory different from Mills's? For Mills, there is intersubjective agreement (though *not* collective acceptance in Searle's sense) about the application conditions for racial

concepts. Their being fulfilled is sufficient for races to exist. For Root, there must be explanatorily significant kinds corresponding to the practices of racial classification. It is not sufficient that there be objective facts concerning who counts as black, white, and so on for there to be races. The mechanism which Root believes produces significant kinds for race is the existence of regulative norms attached to a community's practices of classification. But they would not be socially constructed kinds if they were not also (at least partially) constituted by practices of racial classification.

Haslanger's (2000, 2019) theory of race is somewhat similar to Root's. For Haslanger, a constructionist account "proposes that the conditions for being a member of a racial group are to be given in social terms, rather than in physical, biological, or other non-social terms" (Haslanger 2019 22). Her specific version of this is the following:

> A group G is *racialized* relative to context C iff members of G are (all and only) those:

> (i) 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) …

> (ii) 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

> (iii) whose satisfying (i) and (ii) 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 (i) and (ii) plays a role in that dimension of privilege or subordination. (ibid. 25-6).

The theory identifies races with racialised groups, and individual racial identity with membership of a racialised group. Racialised groups are "demarcated by the geographical associations accompanying perceived body type, when those associations take on evaluative significance … [and] situates the groups on a social hierarchy." (ibid. 26).

What makes this constructionist? Beyond saying that it is a product of social practice, nothing in the account as given actually makes it constructionist in an interesting sense. But Haslanger (2012b) outlines further claims which make it distinctively constructionist. The idea is that what we *take* our race concepts to be tracking (the *manifest* concept) and what they *actually* track (the *operative* concept) diverge. This theory is a specification of the operative concept. The manifest concept is of a group of natural kinds. That's what we take our terms to be tracking. But they actually track social kinds, identified by complex patterns of social relations. These kinds are socially produced, they are social entities. They are the result of - in fact, they just *are* - patterns of human action. The "model" of the relationship between our social practices and the kinds in question is one of *identity*. Being represented as a member of a kind, and thus being marked as fit for a certain social position in the background ideology, *just is* being a member of the kind. The remarkable (and distinctive) thing about covert social constructions is that we take our concepts to be tracking natural kinds, but they are in fact tracking social ones, in part constituted by our practices of representation.

This section has examined four constructionist accounts of race. On Mills's, race is epistemically objective as a result of intersubjective agreement on the membership criteria for races. Racial membership is determined by satisfaction of these criteria. On interactive kind accounts, like Piper's (1992), racial membership is constituted by the causal effects of social practices. Races are groups of individuals upon whom there are such effects. On Root's theory, there is a kind of logical or metaphysical necessity obtaining between our practices and

representations and the existence of race. Races are epistemically thick kinds which wouldn't exist in the absence of such practices. For there to be races, there must be racial classification. On Haslanger's theory, races and individual racial identities are constituted by social relations, making them social kinds. What is special about covert social constructions is that what we take our concepts to be tracking and what they actually track diverge in a special way. We take our concepts to be tracking natural kinds, but they actually track social kinds.

The four accounts show a range of possible models for the construction relation between social factors and races or racial identities. Do any of them vindicate the metaphysics of race debate? In saying that races exist, and that individuals have racial identities, there is a disagreement, seemingly substantive, between them and the racial sceptic. In saying that races are not of independent epistemic interest absent practices of representing them (as Mills does), or that races are defined by the causal effects of our representations (as interactive kind theories do), or that races don't exist independently of our representations of them (as Root does), or that races just are groups positioned in a social hierarchy by our social practices (as Haslanger does), each of these views clashes with biological realism, which holds that races are biological groups. There thus seems to be a substantive disagreement between the parties to the debate (satisfying (1) and (2)), though we will see good reason for doubting this in Chapter 3. Constitutive interpretations are closely related to paradigmatic constructionist theses, since they are either themselves such theses or are interpretations thereof.

1.6. The Constructionist Claim

The best interpretation of the social constructionist claim when made by metaphysicians of race is primarily about what constitutes races and being a member of a race. It is not primarily about what causes or constitutes our racial representations. Nor is it primarily about the causal origins of races or racial membership. While there are disagreements between the parties to the debate concerning some of the causal details, this doesn't exhaust the disagreement, and stems from a logically prior disagreement about what constitutes race. The best interpretation

of the claim is that races and racial identities are *constituted* by human decisions, actions, practices, and institutions. These constitutive accounts, which should not be interpreted as Searlean or Lewisian, are closely related to paradigm social constructionist theses.

This chapter has *not* established that a constitutive interpretation of the constructionist claim does succeed in vindicating the metaphysics of race debate. Chapter 3 will present good reasons for thinking that there is no interpretation which does so. This is because the metaphysics of race debate is non-substantive. What this chapter has argued is that the best hope of vindicating the debate is probably a constitutive interpretation.

Chapter 2: What's the point of a metaphysics of race?

The best interpretation of the constructionist claim in the metaphysics of race debate is that there is a constitutive relation between certain social factors and races and racial identities. Constructionists are best interpreted as saying what it is to be a race or a member of a race. Barnes has argued that analogous approaches in the metaphysics of gender (2020) and disability (2016) are mistaken. This chapter discusses Barnes's position, applying it to race. Barnes's position encompasses two key theses. As applied to race, Barnes's *negative thesis* concerns what the metaphysics of race does *not* have to do. In particular, it need not fit with ordinary racial language and thought. This is explored and defended in section 2.1. Section 2.2 explores whether it has consequences for how we ought to think and speak in terms of race. Barnes's *positive thesis*, explored in section 2.3, concerns what the metaphysics of race ought to do, if not that. Barnes thinks metaphysicians should say what there is which unifies or explains race. I argue that this is not work for metaphysicians - there is no explanatory work left for the metaphysician when the special and social sciences have performed their explanatory work. Section 2.4 explores the metaphysics of race debate more broadly in light of the negative thesis.

2.1. Barnes's Negative Thesis

The negative thesis concerns what the metaphysics of race ought not to do. Barnes's (2020) discussion concerns the metaphysics of gender. She claims that the metaphysics of gender debate is focused on the question of what it is to be a woman (or, generally, a member of a gender). Metaphysicians of gender tend to take providing a metaphysics of gender to at least *essentially involve* saying what it is to be a woman. Sceptics about the metaphysics of gender, who don't think gender admits of a metaphysics, are often guided by the thought that there is no specific thing that it is to be a woman. The debate is (in part) about which theory gets the extension of the ordinary term 'woman' or the ordinary concept *woman* right. Theories are proposed, and counterexamples given, intended to demonstrate extensional inadequacy. Thus, Haslanger's (2012) theory of gender can be criticised for failing to always include transgender

women as women (Jenkins 2016 398), and for counting some transgender men as women. Barnes argues that this is the wrong question to focus on: metaphysicians of gender don't need to be giving theories which neatly map onto, and thereby directly account for, the extension of ordinary language gender terms or ordinary gender concepts, or onto true natural language sentences and ordinary beliefs. Call this kind of mapping *fitting with ordinary language*. The negative thesis (with respect to domain D) says that a successful metaphysics (of D) need not fit with ordinary language (about D).

Barnes notes that many realist metaphysicians hold that metaphysics generally doesn't have to fit with ordinary language: they endorse a general version of the negative thesis. While there is no consensus, this claim enjoys widespread support in the dominant post-Quinean tradition in analytic metaphysics. Sider (2011 171-173), for example, presents a (rather extreme) metametaphysical picture which supports it. The general negative thesis is plausible because it would be surprising if the best theory of the universe could be stated in ordinary language. Contemporary English (and other languages) did not develop to perform that task. Given the realist perspective adopted by most analytic metaphysicians, it just doesn't seem that our metaphysics should be constrained by current, actual habits of speech. Without any reason to think that the domain of race is special, then, this is a good *prima facie* justification that (for realists about metaphysics) one's metaphysics of race need not fit with ordinary racial language.

The characterization of *fitting with ordinary language* remains vague. Consider a specific example. It seems plausible that if an ordinary speaker said, in an ordinary context, 'There are rocks (somewhere in the world)', they would say something true. The claim that metaphysics doesn't have to fit with ordinary language says that accepting the truth of 'There are rocks' doesn't commit one to a metaphysics of rocks. In the metaphysician's ontology, there need be no rocks just because they accept the ordinary English sentence 'There are rocks' as true, and accept that the extension of 'rock' is non-empty. As Sider (2011 13) proposes, a metaphysics can be thought of as an ordered pair $<I, T_I>$, where I is an *ideology* (a set of expressions with associated meanings) and T_I is a theory stated in terms of that ideology. The thought is that

the ideology need not include any expression equivalent to the ordinary English 'rock' or 'is a rock', and there need be no entities in the theory's ontology (if it includes one) of which 'is a rock' is true. What must be the case is that the structure described by the theory be consistent with the truth of the ordinary English sentence 'There are rocks', which must be ultimately explicable in terms of the theory. So, the precise version of the negative thesis (with respect to race) is

> The negative thesis: a successful metaphysics (of race) can differ in ideology from ordinary (racial) language, and its ontology need not include all or only entities in the extension of every natural language (racial) predicate.

Some philosophical accounts tie ordinary language terms to things in the world by mentioning a term on one side of a biconditional and referring to something in the world on the other. They are (often) accounts of the reference of an expression. So, for example, one way of approaching knowledge is by giving an account of the truth values of knowledge ascriptions: "S knows that P' is true just if S and P satisfy conditions $C_1, \ldots, C_n$.' Putnam's (1975) account of natural kind terms (roughly) suggests, for example, that the referent of the ordinary language term 'water' is H_2O. Call these *mention-type analyses*. Other accounts only use, and do not mention, expressions on both sides of the biconditional. So, one could say that water is H_2O, or that knowledge is a propositional attitude obtaining between a subject S and proposition P where conditions $C_1, \ldots, C_n$ obtain. Call these *use-type analyses*. Having a term mentioned on one side of the biconditional allows for the context-sensitivity of an expression to be accounted for. So, if one thinks that 'knows' expresses a different relation in different contexts, then giving a mention-type analysis of knowledge can be advantageous. Analyses which attempt to provide necessary and sufficient conditions are rarely successful, and the search for them has grown less popular in many areas of philosophy. In the case of epistemology, for example, the search for a set of necessary and sufficient conditions on knowledge has been challenged by the persistent recurrence of Gettier-style counterexamples (Williamson 2000 29). It is a consequence of the negative thesis that an account of race need not mention *or* use

ordinary racial terms on either side of the biconditional. It need not be an account of the reference of racial terms, or of the nature of the referents of ordinary racial terms. One can identify metaphysically privileged structures which can then be used to explain racial phenomena, our use of racial terms among them.

Barnes (2020) advocates the negative thesis with respect to gender. It doesn't have to include or neatly match up with true natural language sentences or ordinary beliefs about gender, nor does it have to give the application conditions or extension of ordinary gender terms like 'woman'. Consider an account of gender which does what Barnes says metaphysicians of gender typically, but need not, do. Haslanger's (2000) account says what it is to be a woman. On her account, to have a gender is to occupy a certain social position. A person, S, is a woman just if:

(i) S is regularly and for the most part observed or imagined to have certain bodily features presumed to be evidence of a female's biological role in reproduction;

(ii) that S has these features marks S within the dominant ideology of S's society as someone who ought to occupy certain kinds of social position that are in fact subordinate (and so motivates and justifies S's occupying such a position); and

(iii) the fact that S satisfies (i) and (ii) plays a role in S's systematic subordination, i.e., along some dimension, S's social position is oppressive, and S's satisfying (i) and (ii) plays a role in that dimension of subordination.

A dual definition exists for S's being a man, where S is systematically privileged and perceived to be male. Haslanger's key claims regarding gender are displayed in these definitions. It is an example of a use-type analysis.

Haslanger's account thus presented is vulnerable to objections concerning the extension of 'woman'. First, some men might be mistakenly thought to possess features which are taken as evidence of a female's role in reproduction. Convincing, consistent cross-dressing would be a metaphysical impossibility on this account. Second, transgender women will sometimes be perceived or known not to have such features (Jenkins 2016 400-401). But they are women, and it is important that they be so regarded. As is often the case in debates about the metaphysics of social categories, this brings together normative and metaphysical considerations. But one can make the objection without explicitly bringing in normative considerations. One can say that transgender women are women, and that an account of women which gets this wrong is thereby falsified. The key point is that Haslanger's account as it stands - where it is attempting to fit the metaphysics with ordinary language - is vulnerable to such objections.

Barnes suggests modifying the account. Haslanger's account can be changed so that it doesn't tell us what it is to be a woman: replace 'woman' with 'feminized in a context' and 'man' with 'masculinized in a context' in Haslanger's definitions. Hierarchies structured by feminization/masculinization are the social reality posited by the metaphysics. This theory is correct if all the complex phenomena associated with gender can be explained just with appeal to these structures. The account does *not* say that this is all there is to gender. It says that further components of gender, such as gender identity and gender expression "can ultimately be explained in terms of the basic binary social structure that attributes social significance to perceived biological sex, and which privileges some and disadvantages others based on assumptions about what ought to follow from being … male or female" (Barnes 2020 12). Importantly, since this account doesn't say what it is to be a woman, it is not vulnerable to the counterexamples discussed earlier. The account does not equate being feminized with being a woman, and so it does not matter that some women are not consistently feminized.

There are analogous results applying this approach to race. We can start by attempting to produce a modified Haslangerian theory of race, analogous to Barnes's modified Haslangerian

theory of gender. Recall from Chapter 1 that Haslanger's (2000, 2019) theory of race defines what it is for a group to be *racialised* and for individuals to be members of racialised groups. S is white (in C) just if whites are a racialised group (in C) and S is a member. To be a race is to be a racialised group, to be a member of a race is to be a member of a racialised group. Modifying the theory in the analogous way to Barnes' modified Haslangerian theory of gender leaves the theory as stating what it is to be a *racialised group*, and what it is for an individual to be *racialised* in a context. Races aren't identified with racialised groups, and membership of a race is not identified with membership of a racialised group. The explanatorily important categories are racialised groups and individuals. The other phenomena of race - its importance to a person's identity and culture, for example - are to be explained by appeal to racialisation.

In the case of gender, Barnes argued that the modified version of Haslanger's account avoided some of the objections regarding overinclusion and exclusion. Inclusion and exclusion problems also afflict theories of race, concerning particularly the phenomena of passing and travelling (Mallon 2004). Some philosophers, such as Root (2000 633), think that race is culturally local, that in contexts where a certain racial concept is not used to divide people, people cannot fall under it. So *black* should apply to people only in contexts where the concept is used to divide people. But insofar as such an account is an account of our ordinary concept *black*, this seems to be a flaw in the account, for it may be that our ordinary concept *black* (as employed by us) does apply to people in modally and temporally distant locations where the concept is not used. Passing is an important phenomenon, discussed by Mills (1998). It refers to cases where someone is taken by many, sometimes even themselves, to be of one race when they possess some feature which would usually be sufficient to mark them out as a member of another. This can occur where people's parents are clearly members of one race, but this isn't widely known, and the person's morphology is not that of the stereotypical members of that race. An account which places significant emphasis, as Haslanger's does, on perception in the determination of an individual's race may fail to adequately capture this phenomenon. But these objections do not gain any traction if the accounts are no longer supposed to fit with

ordinary language. So, as in the case of gender, certain objections no longer apply once the negative thesis is accepted.

One could object to Barnes's invocation of realist metaphysicians' support for separating ontology from ordinary language. Those who say that we don't need to have rocks in our ontology just because we admit that 'There are rocks' is true, and that the extension of 'rock' is non-empty, tend to think that this is because 'There are rocks' is made true by the existence of some (relatively) fundamental entities (typically microphysical particles). What we have to include in our *ontology* is just what there is fundamentally, or as Sider slightly differently has it, what there is, in the fundamental sense of 'there is' (2011 6). But it's implausible that racialised groups are fundamental entities - no one thinks they are. On a social constructionist theory, they could not exist without human social practices, so they can't be fundamental - they depend on something else, which itself is not fundamental! Biological categories are also plausibly non-fundamental. But if metaphysics, and ontology in particular, is to include the non-fundamental, the rationale for including racialised groups but not races may seem obscure. The objection is that unless we are restricting our ontology to what there is, fundamentally, we are in need of a principled reason for including some non-fundamental entities (racialised groups) and not others.

Barnes (2020) rejects the idea that broadening the purview of metaphysics beyond the fundamental commits us to a metaphysics that fits with ordinary language. Barnes's point is not that "folk" social categories are non-fundamental and that the divisions into feminised and masculinised (and, analogously, into racialised groups) are. Rather the point is "simply that the most interesting or explanatory categories don't map on neatly to our ordinary language terms and/or common-usage gender categories" (Barnes 2020 23). The metaphysical categories of interest "are slightly different than 'woman', 'man', etc." (ibid. 24). The concepts most useful for understanding gender (or race) do not (or may not) divide the world exactly as do the everyday concepts. 'There are women' is true because there are women, in one sense. But in a different sense, to explain why there are women, we may need to use different

48

concepts. On her proposal, the concepts we need are *masculinised* and *feminised*, as she defines them. These pick out the structures which bear the explanatory weight.

This response explains how a realist can accept the necessity of positing non-fundamental entities without accepting every entity which occurs in the range of a natural language quantifier. Those which are of explanatory importance, those which participate in our best theory of the universe, are those which we require in our metaphysics. One might argue that this just means that they are fundamental. But that would exclude the possibility that explanatorily important categories may be constituted by more fundamental entities. While some might argue for this, it is far from settled. Until such a claim is established, it is legitimate for realists to accept non-fundamental entities in their ontology on the grounds of explanatory power, allowing metaphysicians of race to accept racialised groups, for example, without giving an ontology of races.

In this section, I have laid out the negative thesis as applied to the metaphysics of race. A successful metaphysics of race need not fit with ordinary racial language. Realists should accept this claim, given that they accept it in regard to metaphysics generally, unless there is particular reason to reject it with respect to race. The fact that race is non-fundamental is no reason to reject it.

2.2. The Negative Thesis and Language

The negative thesis implies that a metaphysics of race may have a complex relation to our ordinary race terms. Barnes notes that Haslanger's unmodified theory ties the language to the metaphysics. This allows for a semantic externalism with regard to the meaning of gender terms. They can be given a semantics like the Kripke-Putnam semantics for natural kind terms (Kripke 1980, Putnam 1975). One upshot of this approach is that users of a term don't need to know much about its referent. Putnam (1975) discusses the example of different kinds of tree. I don't need to be able to tell apart a beech from a birch to be able to use 'beech' and

'birch' correctly. Similarly, to use 'woman' correctly, I don't need to *know* that 'woman' applies to those people Haslanger's theory identifies as women. I just need to know that 'woman' applies to women. I don't need to be able to give a use-type analysis of womanhood. The fact that ordinary users don't know that the term refers to feminized people, and may mistakenly believe that the term applies to members of a biological kind, does not thereby show Haslanger's theory to be false. This is important for the constructionist, because if it were not the case that we could (all) have mistaken beliefs about the nature of the referents of our terms, then it could not be the case that, for example, 'woman' applies to a social rather than a biological kind when most people think of women forming a biological kind (Haslanger 2012a).

If, instead, as Barnes suggests, we drop this direct connection between ordinary language terms and the metaphysics, two questions arise. One is whether the metaphysics has any normative implications for how we use race terms. The other is how the extension of race terms is determined, and how the truth-value of sentences about race is determined. In answer to the second of these questions as applied to gender, Barnes adopts a permissive, context-sensitive account of the meaning of gender terms. She assumes a basic level of semantic externalism: she assumes that speaker intention doesn't fully determine speaker meaning, so the two can come apart. She further assumes that there is no single, determinate meaning for each gender term. 'Woman' does not mean exactly one thing. Barnes thinks this is true of most natural language terms, and even more plausibly true of messy social terms like 'woman'. This implies that the extension of 'woman' varies. The exact mechanism for the variation in meaning could be a number of things. Most obviously, it could be a context-sensitive expression or it could express an interest-relative property.

An objection might arise to this permissive semantics. Barnes assumes a degree of semantic externalism with respect to gender terms. She claims that there are metaphysically privileged structures closely related to the meaning of gender terms. Won't these metaphysically privileged structures act as "reference magnets", meaning that 'woman' will just track

feminised persons? Barnes thinks not. She thinks that reference magnets are typically thought to "*sometimes* trump use, or [to] be a tiebreaker when use is ambiguous" (2020 24 n.45). And she thinks that our use is sufficiently variable and distant from the Haslangerian structures that the mere existence of such structures shouldn't trump that usage. One might add that just because one adopts semantic externalism with respect to a term - particularly the attenuated version that Barnes assumes - doesn't mean that one is committed to the doctrine of reference magnetism at all. To adopt semantic externalism is just to say that the meaning of an expression depends in part on factors external to the speaker. This can include patterns of usage by other speakers. It doesn't commit one to any claims about the existence of reference magnets. And as Barnes suggests, even if one accepts some version of the doctrine of reference magnetism, one can reject the idea that reference magnets *always* trump use. Indeed, one should, for otherwise one runs the risk of disallowing important cases of empty reference such as 'phlogiston'. All one has to say, therefore, is that Haslangerian social structures (or whatever structures one's metaphysics posits) are not sufficiently strong reference magnets, or are not sufficiently close to the referents of ordinary social category terms, to outweigh ordinary usage.

What about the question of whether the metaphysics of race has implications for how we ought to use race terms? Barnes claims that realists who accept the general negative thesis are likely to view the meaning of our ordinary language as largely determined by use. Whatever the true metaphysical theory of the universe, there are rocks. Our use of 'rock' more or less determines its extension. So, presumably, our metaphysics of race isn't going to have revisionary implications for how we use race terms. The influence of use on meaning means that 'x is black' is true in a context just if our ordinary usage of 'black' in that context would apply to x. The idea is that ordinary language can't be systematically mistaken. While ordinary language users can be mistaken in applying a term like 'rock' in a given instance, for example when they mistake something else for a rock, it could not be the case that we are all collectively mistaken about what rocks are. How we use 'rock' determines the meaning of 'rock', it determines what 'rock' applies to. Barnes thinks that the social constructionist will want to

resist this because they have critical ambitions. They think that there is something systematically wrong in our use of gender (and race) terms.

Barnes attempts to solve this problem by identifying some revisionary implications that a metaphysics can have for how we use gender terms. Unlike for 'rock', people *can* be systematically mistaken about gender. So, a metaphysics of gender can have implications for how gender terms should be used. Why is this possible for gender terms but not terms like 'rock'? And how do these implications arise? Barnes thinks that the metaphysical commitments of 'There are rocks' are far more remote than the metaphysical commitments of 'There are women'. For gender, the metaphysical structure is a closely related category to the extension of 'woman' in most contexts. For chairs, it is microphysical particles. That's the principled explanation of why ordinary gender terms are supposedly responsive to the underlying metaphysics. But because Barnes has said that the truth-values of ordinary gendered sentences and the extensions of gender terms don't neatly map onto the social metaphysics, it's difficult to understand *how* the metaphysics is to exert this influence. Part of the point, for Barnes, was to say that the metaphysics shouldn't commit Haslanger to a certain, context-invariant, extension of 'woman', and hence to certain truth-values for 'x is a woman'.

A specific implication Barnes identifies from adopting her approach is that we can let normative considerations guide what we let gender terms mean in a context. Because the metaphysics needn't tell us what it is to be a woman, if one adopts an account where it does *not* and on which 'woman' can latch onto all sorts of different properties and property clusters, there will be a way of interpreting 'woman' in a context which allows us, for example, to vindicate self-ascriptions of gender. This lets us meet the important goal of affirming people's gender identity. Thus, the implication that the metaphysics has comes from the negative claim that there is nothing which it is *really* to be a woman, there is no metaphysically privileged class denoted by 'woman'. One who shares this goal, and who sees it as an advantage for a metaphysics that it can help achieve it, could thus seemingly find Barnes's approach attractive.

There is a problem with this, revealing the limitations on the revisionary implications of the metaphysics. It seems probable, given Barnes's picture, that in some contexts 'woman' can be made to track biological sex or perceived biological sex. In the first case, trans women would not count as women in that context. In the second, some trans women (and others who don't "look" female) would not count as women, and some men, including trans men, would. Thus, while the account *allows* one to meet the normatively important goal of validating people's self-ascriptions of gender, it also allows people to invalidate such ascriptions. Barnes's response to this, which she thinks draws out another way in which the metaphysics has implications for the use of the relevant terms, is that we can use the terms in ways which communicate (presumably through Gricean implicature) claims which are inconsistent with the underlying metaphysics. In that case, what is said, though possibly literally true, is not assertable. If someone says, for example, 'Trans women aren't real women', this, according to Barnes, communicates other claims, like the claim that gender is determined by biology or that there's something wrong with people whose gender identity differs from the gender assigned at birth. These claims are false in all contexts because, Barnes claims, they misdescribe the basic reality of gender. Whether biology determines the complex phenomena of gender is not a claim which is true in some contexts but false in others. It's false in all contexts because it's inconsistent with the (assumed) metaphysics of gender. So, the proposed social constructionist metaphysics does allow the constructionist to justify certain claims about how we ought to use gender (and presumably race) terms.

This solution doesn't work. First, while it might be conversationally inappropriate to say 'Trans women are not women' in any context, it doesn't follow that it is false to believe the proposition it expresses. This is analogous to Edgington's (1995 245) argument against a defence of the material conditional semantics for 'if' based on the pragmatics of assertion. In some contexts, according to the above argument, trans women won't be in the extension of 'woman', so one could truly believe of trans women that they aren't women, without believing anything inconsistent with the metaphysics. Second, it is unclear that the claims Barnes identifies really are inconsistent with the metaphysics she proposes. Barnes uses 'gender' in a

specific, technical way. But it's plausible that there's an ordinary use of 'gender' where it means either being a man or being a woman (and, on this usage, gender *is* binary). Someone who said 'Gender is biologically determined', and meant by this simply that whether one is a man or a woman is determined by biology would be *right* in just those contexts where the extension of these terms track biological kinds. Barnes would want to object that what the sentence 'Trans women are not women' implicates is that what gender *really is* is biological sex. And, although there is nothing which it *really is* to be a woman, there is something which gender *really is*. It's just false that the various phenomena associated (by everybody) with gender are determined biologically, and that's what's implicated. This seems more plausible but is hardly conclusive. After all, one can normally cancel implicatures. So if one said 'Trans women are not women, though of course gender is not determined biologically and there is nothing wrong with people who claim to be trans women and …' one should, on Barnes's account, have said nothing objectionable. But many feminists would not be satisfied with this. In which case, the metaphysics won't have the implications for our gender term usage that Barnes and others want it to.

Let's return to the question of whether the metaphysics of race can have implications for how we use race terms. It is worth getting clearer on what ordinary speakers can and can't be mistaken about. It's plausible that, barring extreme sceptical scenarios, we all know that there are rocks. In ordinary circumstances, ordinary people are reliable judges of whether something is a rock. And our usage determines what 'rock' applies to, and we couldn't be systematically mistaken in applying 'rock'. But we can all be wrong about what rocks *really are*. We can have false beliefs about them. Before the widespread adoption of atomic theory, it may be that everyone thought of rocks as solid blocks with no spaces in them. But really rocks are atoms arranged in certain ways (or something like that). Similarly, one might think, while we can't be systematically wrong in applying 'woman', for the same reasons, we can be wrong in individual cases, and as individuals. And, more importantly, we can be wrong about what the kind 'woman' tracks is like. This is a point stressed by Haslanger (2012 16), and is consistent with Barnes's picture of the metaphysics of gender. We can think 'woman' applies to people sharing

a certain biological feature, when it actually applies to members of a particular social group. This is what makes these categories covert constructions. The error is not in our applying the concept, but in our thinking that we are tracking one kind of structure when in fact we are tracking another.

Is it a problem if the metaphysics doesn't have the implications for our use of social category terms which Barnes wants it to? Not obviously, at least in the case of race. The intuitions around passing and travelling are unclear, and whether we should validate "transracial" self-ascriptions is less well understood. If the correct metaphysics is one which doesn't entail that we ought to use terms in one way rather than another, this means one can argue that certain claims about how we ought to use terms are *not* justified by the metaphysics. Thus, claims about how we ought to use our terms lack justification in the metaphysics. Our usage can be sensitive to a range of normative considerations. This negative implication, which Barnes (2020) does not focus on, is a key upshot of this approach. This issue will be taken up in Chapter 3. We can let the debate about language be an explicitly normative and practical one. Whether or not to go on using racial language, and how to use it if we do, can be divorced from the debate about the underlying metaphysics, to a significant extent.

It follows from the negative thesis that a successful metaphysics of race will not necessarily have normative implications for how we use racial terms. As I shall argue later (3.6), this is a welcome implication of the negative thesis.

2.3. No Work for Metaphysicians

If the role of metaphysicians is not to provide a metaphysics which explains ordinary terms, concepts, true beliefs and true sentences by fitting with ordinary language, what is it? This section examines Barnes's *positive thesis* as applied to race:

Barnes's positive thesis (applied to race): the task of a metaphysics of race is to *explain race*.

This notion of explaining race is obscure. I will argue that there is no distinctive explanatory task for metaphysicians of race to undertake. Instead, our metaphysics of race, being a specification of the categories which allow us to explain the phenomena which are generally thought of as race-related, will be settled by work in other disciplines than philosophy.

Modified positive thesis: the metaphysics of race will specify the structures of significance for explaining the phenomena associated with race, but this will not be a task for the metaphysician.

Barnes says that a metaphysics of gender is supposed to explain gender. What is it to explain gender? Barnes takes it to be:

to say what it is in virtue of which people have genders, or in virtue of which members of a given gender can be said to have something in common with each other, or in virtue of which gendered norms and roles have the significance they do. ... [W]e're trying to say what feature(s) of the world - if any - unify or explain gender. (Barnes 2020 2)

By analogy, metaphysicians of race are trying to identify what unifies race, what explains individuals' having racial identities, and what gives race the social significance it has. And, although this presumably *could* be achieved by saying what it is to be a member of a race, it need not be.

This is still unclear. What is it to "unify" race? How is this distinct from offering necessary and sufficient conditions on, say, being black? Perhaps by considering Barnes's proposed account of gender, we can get a better understanding of what it is supposed to achieve. On

Barnes's modified version of Haslanger's theory of gender, the complex phenomena associated with gender can be explained using the framework of masculinisation and feminisation. On the first gloss on 'explaining gender', Barnes's idea is presumably that it's true to say that women have something in common because they tend to be feminised, or are feminised in certain contexts, or bear some one of a variety of relationships to being feminised. But it's still unclear how this is distinct from giving necessary and sufficient conditions for category membership. It's tempting to think, without a better explanation of what's going on, that successfully pinning this down would at least *entail* a set of necessary and sufficient conditions for category membership.

What about the other glosses? It seems to pick out a *prima facie* plausible candidate for what it is in virtue of which people have genders - it is, at bottom, in virtue of being masculinised or feminised in various contexts - and for what is in virtue of which gender has social significance. The actual explanation is not given, however. It is merely said to exist and make reference to masculinisation/feminisation. A genuine explanation would spell out *how* being masculinised or feminised explains these things. And it is tempting to think that whatever that explanation would be, it is the task, not of the metaphysician, but of the sociologist, social scientist, or psychologist to provide it.

One might argue that this isn't work for the metaphysician because it does not concern the fundamental. But that would not be a good argument. We would require a convincing argument for the claim that metaphysicians should be solely concerned with the fundamental. As Bennett points out, metaphysicians spend a lot of time talking about the non-fundamental, discussing the status of the special sciences, the nature of secondary properties, possible worlds, and works of art (2017 232-3).

A better argument is rather that whatever metaphysicians' work is, it isn't identifying the most explanatorily useful concepts and categories within other disciplines, and *this* is what Barnes appears to be doing. It's up to physicists, for example, to decide which concepts to use to

understand the physical universe. The metaphysician might have something to say about what *in general* the difference is between useful concepts and useless ones. But they aren't there to adjudicate specific questions in other disciplines. There are, of course, seemingly distinctive special questions in metaphysics, like the existence of god or free will or the nature of personal identity. In those cases, it does seem like metaphysicians can decide on the most useful concepts and categories, because these topics are plausibly the exclusive purview of metaphysicians. But the task of understanding race seems different from these. It is plausibly a sociologist's or a social scientist's or a psychologist's. We are in need of a justification for philosophers, and especially for metaphysicians, to intrude here. My major departure from Barnes is in holding that there is no such justification - there is no work for metaphysicians of race.

One response to this is to say that the boundaries between disciplines are porous and fuzzy. Philosophers of physics and biology do research which is quite close to theoretical physics and evolutionary theory (Williamson 2007 18). To police these boundaries might foreclose interesting research, and prevent helpful inter-disciplinary engagement. This is correct insofar as it goes. But it fails to identify a role for metaphysicians, merely insisting that there might be one. Perhaps philosophers can helpfully clarify concepts and some of the points at issue in wider debates about race. But then metaphysicians are employing their skills in order to assist those working in other fields. They are not themselves doing the metaphysical work. A distinctive task needs to be identified for metaphysicians to undertake.

One might reply that what metaphysics is is partly determined by what metaphysicians *do* when they are doing (what they take to be) metaphysics. But it isn't completely determined by this. A confused metaphysician who spent their time proving results in number theory thinking that they were doing metaphysics would not be correct. One might insist that it is at least much less likely, if not actually impossible, that serious metaphysicians should be systematically mistaken in thinking that what they are doing is metaphysics. But if so, it should be possible

to point to the distinctive task they are engaging in, and to explain what that task is and why it is a task for them.

One might instead just shrug at the charge that this is not for metaphysicians. Perhaps it isn't part of the same discipline which engages in attempting to answer questions like 'what is there?' and 'does free will exist?' and 'what is a cause?'. If philosophers can contribute to debates in other fields, then that's all to the good. But to adopt this response is to accept that the substantive work is not for metaphysicians. While philosophers may perform useful work, such as showing that social constructionism is compatible with naturalism and realism (Mallon 2007a), this is presumably all that they contribute. Identifying which categories are of explanatory importance and providing the explanation is a matter for other disciplines.

What is needed is an explanation of what the distinctive metaphysical task *is*. Consider another metaphysical debate: the debate about the metaphysics of musical works. This is a debate conducted by metaphysicians. Some people think they're sets of performances, others that they're scores. Others still think they're types of sound sequence, and some think they're nothing at all. The debate is not a debate between musicologists. Metaphysicians think there are various truths that need explaining concerning musical works. For example, Dodd (2007) gives several features of musical works that require explanation: one can hear them; they are repeatable; they are composed. It is perhaps difficult to articulate exactly *why* explaining this is a task for the metaphysician. But it plausibly is one. Of course, there *are* similar-sounding questions which a musicologist or a biologist or a physicist would be the right person to answer. But if we knew everything about the physics, biology, and musicology, we would still have the distinctively metaphysical questions remaining. The nature and reality of musical works needs pinning down. The underlying reality must be able to explain, for example, how we can hear musical works. By analogy, if we knew all the sociology, the biology, and the social science, Barnes must insist that there would be an *explanatory remainder*: something left for the metaphysician to explain.

This is the crux of the matter: what is the explanatory remainder left after sociologists, biologists, and social scientists have done their work? In particular, if it's not specifying necessary and sufficient conditions, what is it? Barnes's glosses are insufficient to explain what the metaphysics is supposed to be doing. In fact, they are confusing. On one of them, to explain gender is to say what it is in virtue of which members of a given gender can be said to have something in common with each other. As noted earlier, if there is something in virtue of which members of a gender can be said to have something in common with each other, this would seem to imply that there is a set of necessary and sufficient conditions on being a member of a given gender.

In Barnes's first gloss on 'explaining gender' - theorising what it is in virtue of which people have genders - there is an ambiguity. On one reading, this is theorising what makes an individual have the gender they have. So, it is explaining what it is in virtue of which that I am a man, that Barnes is a woman, and so on. This sounds very much like explaining what it is to be a man or a woman or non-binary. On the other reading of this gloss, it is explaining why it is the case that anyone at all has a gender. What is it about the world that makes it the case that people have genders? These both seem like important explanatory projects. But neither, nor the analogues for race, are tasks for metaphysicians. Between them, the sociologist, psychologist, and biologist will surely provide an adequate explanation of what being a woman consists in, as well as a causal explanation of how one comes to be a woman, and hence of what about the world makes it the case that people have genders. Unless the question is about the application conditions of ordinary language terms, then there seems no work for a philosopher here. And the task of giving application conditions for ordinary language terms is not the primary task of metaphysicians, but of philosophers of language and semanticists.

Perhaps we should depart from Barnes's glosses on explaining gender. Perhaps the explanatory remained (for race) is the following: There are various racial systems. Race is culturally local. There is something in common between, say, being black in New York and being black in London. There is also something in common between, say, being black in New York and

being white in Paris. There is *also* something in common between races, and something in common between racial systems. The task left to the metaphysician is to specify what these various commonalities consist in, which may end up being the project of giving necessary and sufficient conditions for being a racial system or for being a racial identity. The analogous answer to what Barnes advocates for being a woman in the UK and being a woman in the US, and for genders, and for gender systems, is that being black in New York and being black in London share a certain relation to practices of racialisation, that the category of black people in the US and white people in the UK share the fact that their members are routinely racialised, or that being a racial system is to be a racialising system. This doesn't imply a particular extension for *black*, *white*, or *race*. But the objection remains that this is not plausibly a task for the philosopher. Sociologists and psychologists will develop concepts to divide up the terrain in certain ways in order to explain certain facts. In so doing, they will pick out these categories which have something in common. If, in order to understand various societies, one can and must use a concept of racialisation, then a successful sociology will employ this concept. It is not up to the philosopher to identify the most explanatorily useful category unless there is a distinctively philosophical explanatory remainder. And this doesn't isn't such an explanatory remainder.

Mallon (2016, 2017) suggests that the causal efficacy of social kinds is an important feature in need of explanation. If this proved to be a distinctively metaphysical explanandum, then it would constitute an explanatory remainder. The particular fact that needs explaining is that race features in stable explanatory and predictive generalizations. Positions in the metaphysics of race as standardly interpreted give one set of possible such explanations. The biological realist will say that being a member of a race is causally efficacious because of the same kind of mechanism which makes being a member of a species causally efficacious. The racial sceptic will deny that being a member of a race is genuinely causally efficacious, and that we need to posit races to explain the successful generalizations which can be made. The constructionist says that races are causally efficacious because of a mechanism which characterises them as social kinds (Mallon 2017). But once again this would appear to be a matter to be decided by

investigation in the relevant special or social science. Whether the causal effects of race are sustained by a biological mechanism or by a mechanism which characterizes races as social kinds depends on empirical issues. Unless the issue is whether ordinary racial terms apply to kinds discovered by the social and special sciences, which is not a characteristically metaphysical question, it is unclear what the philosopher's task is.

It is the role of philosophers of language and semanticists to identify semantic relations between expressions and their referents (assuming a referential theory of meaning). So it is, in a sense, for philosophers to discover what the referent of 'water' is, and how that is determined. But their role is limited to identifying semantic relations. It is not for them to discover the true nature of the referents. Thus, a philosopher who has an account of the semantics of 'water' where it refers to that stuff in the world, water, and in which its contribution to the truth conditions (or, more generally, the meaning) of a sentence in which it takes part, has a theory of 'water', and need not investigate the nature of water. On the other hand, an empirical scientist *does* need to investigate what water is, what its properties are, and so on. The natural scientist doesn't need a theory of the meaning of 'water'. They need a theory of water. The role of the philosopher in finding out what water is is therefore limited, debates about whether water (necessarily) is H_2O notwithstanding. By analogy, the philosopher has a role in investigating ordinary racial thought and talk. In particular, the messy and apparent context-sensitivity of gender and race terms is of interest, as are the pragmatic implicatures and other significant communication systems around race. But philosophers don't seem to have a role in identifying what the referents of our ordinary racial terms and concepts *really* are. And if "explaining race" does not even involve talking about the referents of ordinary language, as per the negative thesis with respect to race, then it is even more obscure what the philosopher's role would be.

In this section, I've argued that there is no work for metaphysicians of race. Once we adopt the negative thesis, we abandon the view that a metaphysics of race has to fit with ordinary racial language. We are left with the task of saying what is left for the metaphysician of race to

do. They are not in the business of conceptual and semantic analysis. So, it cannot just be analyzing ordinary racial thought and talk. But there is no distinctive explanatory task for the metaphysician. If our metaphysics of race is our best overall theory which explains the various phenomena of race, it will be delineated by other disciplines. We should reject Barnes's positive thesis (insofar as it is committed to there being a distinctive explanation for metaphysicians to provide). Instead, we should adopt the *modified positive thesis* that the correct metaphysics of race is a specification of the categories which *explain race*, which are those which are necessary in understanding the phenomena associated with race, but that this work is not work for metaphysicians of race.

2.4. Biological Realism, Racial Scepticism, Racial Pluralism

This section will explore the impact of adopting the negative thesis and the modified positive thesis on the other positions in the metaphysics of race debate. Note that while it has been argued that these positions aren't to be established by metaphysicians, it is still of interest to consider what claims about race amount to in light of the negative thesis - that the metaphysics of race need not fit with ordinary racial language - and the modified positive thesis - that the metaphysics of race amounts to a specification of the explanatorily significant categories identified by other disciplines.

Once we adopt the negative thesis and the modified positive thesis, the three major positions amount to claims about the explanation of race. Biological realism says that the explanation of the phenomena associated with race is biological. Racial constructionism says that these phenomena are to be explained by appeal to something social, in particular certain human decisions, actions, and institutions. Racial scepticism says that there is just nothing which explains or unifies these various phenomena. Once we adopt the modified positive thesis, it becomes clear that which of these positions is correct depends on which fields are able to explain which phenomena.

2.4.1. Biological Theories of Race

On the standard interpretation of the metaphysics of race debate, explored in Chapter 1, the biological realist says that membership of a race is a biological property, and that races are biological categories of some kind. An objection to a biological theory of race thus interpreted would be that it gets the extension of our race terms wrong. So, for example, Kitcher's (1999) theory can be faulted for counting the Amish as a race, but not white and black people in the US. Such objections aren't insurmountable. One could, as Andreasen (2005) does, adopt an ambivalence about whether races *still* exist. Or one could, as Andreasen (2005) also does, insist that a certain deviation from folk theories is acceptable if the categories identified are useful for biologists. If we adopt the negative and modified positive theses, this second response is stronger. It's not a problem for the biological realist if the extensions of their race concepts aren't the same as the "folk" races, or if the biological races are different from "folk" races. Call the groups identified by the biological theorist *racial sub-populations*. Then, according to the biological theorist, we can explain race in terms of racial sub-populations. To explain race, we don't need something biological, not social, in our metaphysics.

Some of the objections to biological theories of race don't apply once we adopt the two theses. But some other, weighty, objections *do*. It would appear hopeless to explain race purely in terms of racial sub-populations. For example, take the fact that one's race is *socially significant*. Being a member of one race rather than another significantly affects one's life prospects. But it looks hopeless to explain this in terms of being a member of a racial sub-population. Whether it is is not a matter for metaphysicians to rule on. But unless it turns out that these differences in socially significant properties can be explained in terms of membership in biological categories, biological realism looks false.

What makes this prospect look particularly hopeless is that some biological realists' posited racial categories don't come close to the ordinary divisions between, say, white and black people in the UK. But these groups have systematic differences which need explaining. Take

the Kitcher-Andreasen thesis on which races are relatively reproductively isolated sub-populations. Reinterpreted in light of the two theses, the claim is that race is explicable by reference to these reproductively isolated sub-populations. One is not committed to the claim that these just *are* races, or that racial identity is just membership of one of these groups. But white and black people in the UK just aren't very reproductively isolated. Almost one in forty people in England and Wales were mixed race or reported multiple ethnicities in the 2011 census (ONS 2012). This is more than 10% of the non-white population. This is plausibly enough to undermine the possibility of dividing people up into reproductively isolated populations which roughly match up to white and black "races". So, this theory will very probably lack the resources to explain various racial phenomena.

Even a biological realism which explicitly sets out to map onto ordinary racial categories, such as Spencer's (2019a), has trouble. Spencer's claim is that successful biological theories posit racial categories which ordinary race terms track. Even if Spencer is right that biological theories posit racial categories which are the referents of ordinary racial concepts and terms, it still doesn't look hopeful that race can be explained by biological theory. It seems improbable that it will be able to explain social significance. Consider that people who lack whatever underlying biological features which the biological theory tracks using the concept *black* could still be perceived to have African ancestry and, on the basis of this, be treated as black. It seems likely that such a person would experience significant oppression. But there are no resources in the biological theory on its own for explaining this. This is not the objection that the biological theory gets the extension *wrong*. That would not be a valid objection for the biological theory interpreted in light of the two theses. Rather it is the objection that Spencer's theory is explanatorily inadequate.

2.4.2. Racial Scepticism

How should racial scepticism be reinterpreted in light of the two theses? As standardly interpreted, the racial sceptic denies the reality of race. In light of the two theses, the sceptical

position should instead be seen as the view that there is nothing which explains or unifies the phenomena of race. No metaphysics of race can be given. Race is just a jumble of inconsistent and incoherent practices and beliefs. There is no explanation of why people have races, no explanation of race's significance or causal power.

Racial scepticism thus interpreted is more extreme than on the standard interpretation. Appiah's view (1996), for example, that racialisation explains the social significance of race is better seen as a constructionist view than a scepticism, on this proposal. Glasgow's (2019) view that there are racialised groups which are not races similarly amounts to a constructionist view on this interpretation. Both Glasgow and Appiah admit the existence the constructionist's key explanatory categories. They admit that there are genuine explananda and that these categories can perform the explanatory work required of them. What they deny is that those categories are races. Indeed, as will be explored in Chapter 3, few would deny the existence of racialised groups, where 'racialised group' applies to a group whose members are identified as having certain ancestry based on the (perceived) possession of certain features, mainly morphological, and to which social significance is attached. Furthermore, few would deny that racialisation is at least a major part of the explanation of differences in life prospects. Most parties to the metaphysics of race debate accept that there is systematic racism, and that it is necessary to understand this to understand social life. But a "racial sceptic" who acknowledges the central explanatory importance of structural oppression based on perceived ancestry in understanding race, and in understanding society and social life, is best interpreted as a constructionist on the debate interpreted in light of the two theses. The constructionist no longer has to claim that racialised groups are races, or that being racialised just is being a member of a race. So, it's not enough to reject social constructionism simply to say that there are racialised groups but that these are not the referents of ordinary racial terms and concepts.

This does not imply that racial scepticism is indefensible on the debate so construed. Barnes considers versions of scepticism about the metaphysics of gender which can carry over:

Perhaps what we call 'gender' is nothing more than a loose collection of words and beliefs, or perhaps gender is not any one thing, but rather how we choose to 'perform' in response to sex-based stereotypes, or perhaps it's just implausible that there's any sense of gender that is stable across differences in race, class, sexuality, and culture. (Barnes 2020 2)

The first view is associated with those who prioritize fundamentality in metaphysics, but also certain gender eliminativists. Something like the second is advocated in Butler (1990). The third is advocated in Spelman (1988).

2.4.3. Pluralism

The positions considered so far have in common that they are *monistic*. In the debate as standardly interpreted, biological realism says that race is biological. Races are biological kinds; racial identity is membership of these kinds. Social constructionism says that race is social. Races are social kinds; racial identity is membership of these kinds. Racial sceptics say that race doesn't exist. On the debate interpreted in light of the two theses, biological realism says that the explanation of race is biological. The explanatory categories are biological. The constructionist says that the explanation of race is social. The explanatory categories are social. The racial sceptic says that there is no explanation, no privileged categories to do the explanation.

All of these positions posit (at most) one kind of category, which either *is* or *explains* race. A possibility I've ignored so far is that there is more than one kind of category. Race may be more than one thing, or its explanation may require more than one kind of category.

Consider Spencer's (2019b) proposal. Departing from his (2019a), Spencer argues for *radical racial pluralism*, the view that there is a "plurality of natures and realities for race in the relevant linguistic context" (Spencer 2019b 27). What this means is that there are various answers to

67

the question 'What is race?', asked about a given linguistic context. By 'linguistic context', Spencer means something quite broad, namely a community's entire *race talk* - all of the dominant ways that 'race' and race terms are used to classify people in that community using that community's language(s). Spencer says that those involved in the metaphysics of race debate are charitably interpreted as asking what race is and whether it is real, as the term 'race' and other racial terms are used in contemporary American English. Thus, they are engaged in the *US race debate*, focusing on US race talk. The theories are US race theories, theories about the nature of race in the context of the US.

Spencer seeks to establish radical racial pluralism with respect to the US. That is, he seeks to show that there is a plurality of natures for race in the US. The correct US race theory - the correct theory of the nature and reality of race in the US - is not monistic, but rather pluralistic. It will be a disjunction, where each disjunct is a statement about what race is and whether it's real in a specific context (or part) of US race talk. A monistic theory says that race is one thing. So Haslanger's (2012) theory is a monistic theory, despite positing a context-sensitive property as the referent of race terms. On Haslanger's theory, racialised groups are races, and being a member of a racialised group is having that racial identity. Which groups are racialised, and how, depends on the context. But race is identified with a single, context-sensitive property. A pluralist theory, by contrast, says that race is *more than one thing*. To be pluralist, one can conjoin another monistic US race theory, and reduce each theory's scope to a specific part of US race talk (Spencer 2019b). Consider Andreasen's (2000) theory on which races are clades, reproductively isolated monophyletic groups. A theory which says that races are racialised groups in context C_1 of US race talk or clades in contexts C_2 of US race talk, where C_1 is not the same as C_2, and clades are not racialised groups is a pluralist theory.

Spencer's strategy for establishing radical racial pluralism is to argue for a pluralist US theory of race, which implies radical racial pluralism with respect to the US, and hence establishes radical racial pluralism, generally. His argument is the following:

(1) Radical racial pluralism is true for US race talk if the correct US race theory is radically pluralist.

(2) The correct US race theory is radically pluralist if more than one distinct meaning of 'race' is used in US race talk.

(3) One meaning of 'race' used in US race talk is the OMB's meaning of 'race'.

(4) The OMB's meaning is the set of human continental populations, and the OMB's meanings for 'American Indian', 'Asian', 'Black', 'Pacific Islander', and 'White' are Native American, East Asian, African, Oceanian, and Caucasian, respectively.

(5) The OMB's meaning of 'race' is not the only meaning of 'race' used in US race talk.

(6) So radical racial pluralism is true for US race talk.

Spencer holds that we are seeking a theory that makes sense of, by fitting with, US race *talk*. He thus ties the metaphysics of race with the task of accounting for such talk. So, he ends up with a pluralist theory of race, where what race is varies from context to context. The metaphysics lines up with the extensions of ordinary race talk, but because that race talk has to be divided into different contexts to be made coherent, the metaphysics follows suit. So, we get a disjunctive metaphysics of race because the task of explaining race is taken to be explaining the extensions of ordinary race talk.

If the two theses are correct, then Spencer's strategy is misconceived. Just because race terms vary in meaning from context to context, this doesn't mean that we have to accept a pluralist metaphysics. Compare this to the debate around the context-sensitivity of knowledge ascriptions. DeRose (1995) argues that 'knows' is context-sensitive: 'S knows that P' can vary in truth-value for a given subject S and a given proposition P, while only the context in which the sentence is uttered changes. Suppose this is true. One way of reacting to this would be to say that knowledge is a different thing in different contexts, or rather that knowledge is a plurality of things. In context C_1, knowledge is the property expressed by 'knows' in C_1, and in context C_2, knowledge is the property expressed by 'knows' in C_2, and so on. This is analogous to Spencer's approach.

To account for much the same data, one could instead say that 'knows' expresses a property which itself depends on contextual parameters. This approach is represented by subject-sensitive invariantism, advocated by Hawthorne (2004) and Stanley (2005). On this view, a knowledge ascription expresses the same proposition when uttered in different contexts about the same subject S and proposition P. What varies is whether the person knows, because in some situations knowledge requires more evidence than in others. This is analogous to Haslanger's theory. 'Black', for example, expresses the same property independent of the context of use. What varies is which properties are associated with which continental ancestries, the normative significance attached to this, and whether or not a given individual is perceived to have those properties.

So, one kind of argument for pluralism is undermined by adopting the negative thesis. In fact, a pluralism like Spencer's comes close to a denial of the possibility of a metaphysics of race *simpliciter*, that is, to racial scepticism. If there is nothing but a kind of disjunctive unity (something is a race just if it is F in context C_1, G in context C_2, etc.), with nothing uniting the disparate instances of race other than our habit of using the same terms, there may be no metaphysics of race. As Barnes (2016) says in her discussion of the metaphysics of disability, we're "not simply asking which things count" as races (Barnes 2016 12). If race in one context seemingly has nothing in common with race in another context, a successful metaphysics of race has to tell us what unifies the disparate cases. A disjunctive account doesn't actually point to anything unifying. Someone with Spencer's commitments might choose to adopt a racial scepticism in the debate when it is interpreted in light of the two theses. They might think that there just is *no* metaphysics of race.

In fact, the evidence Spencer draws on to demonstrate different US race talk contexts could be used to support scepticism. Race talk covers many complex phenomena, so perhaps it is highly unlikely that any theory can unify these phenomena. It is hopeless to look for a theory which will explain race in contexts where racial terms apparently track ancestry, perceived ancestry, certain biological features, and so on. Spencer even says that no *global* theory of race

is possible. Race is not the same thing in the US as in Brazil. Spencer doesn't explain why one can have a pluralism with respect to race in the US but not with respect to the globe. But the point is that even Spencer supports a moderate metaphysical scepticism about race, whether or not this sits well with his pluralism about race in the US.

Adopting the two theses does not in fact undermine pluralism, or make it collapse into scepticism. Different parties to the race debate often focus on different aspects of race which require explanation. For example, Haslanger (2000) takes understanding of the mechanisms of oppression as her key explanatory target. Spencer (2019a), meanwhile, takes certain biomedical propensities to be important. So, a metaphysics of race will plausibly have to explain how being a member of a certain race is linked to being oppressed or privileged, and how it is linked to being more or less likely to develop certain diseases. An attractively simple theory would be one on which our race terms just track a certain kind, and to be a member of this kind is to have certain properties which explain *both* being oppressed and being more likely to develop certain diseases. Perhaps the term tracks a social kind, as on Haslanger's (2012) proposal, and being oppressed leads to one developing certain diseases, as Root (2000) suggests. Or the implausible claim that the terms track a biological kind, and there is a biological explanation for the oppression and privilege, might turn out right.

A more complex theory has our terms track different things in different contexts. So, in one context, as Haslanger claims, our terms track perceived ancestry, but in another, as Spencer claims, they track biologically real kinds. In that case, if we adopt Barnes's approach, a metaphysics of race should explain what the races have in common. What is it that *being black* in one context has in common with *being black* in another? The metaphysics should explain this, according to Barnes. But one can reject this aspect of Barnes's approach. The options are more extensive than a monist metaphysics or a metaphysical scepticism. There may, that is, be a pluralist metaphysics of race. By separating the explananda into different clusters, we may find that there are two or more theories which explain and unify separable aspects of race. To illustrate, suppose we divide up the phenomena of race into purely biomedical and purely

social (and assume for now that there is no remainder and no overlap). It may still be that there is significant variation between contexts in the extension of race terms within uses which track social features and those which track biomedical features. If not, that's good reason to adopt the view that there are two kinds in the vicinity, which our race terms ambiguously track. But if so, we can still say that the explanatory categories of interest don't straightforwardly line up with the extensions of our race terms and concepts. There may still be exactly two metaphysically interesting structures, but our ordinary terms and thoughts don't latch directly onto them, at least not always. There is no *one* thing - no one structure or kind - which can explain *all* the phenomena associated with race, no matter how complicated the theory. But if we divide up the phenomena suitably, then we can give two theories which explain two clusters of phenomena.

2.5. Race and Metaphysics

In this chapter, I've explored the negative thesis with respect to race. A successful metaphysics of race does not need to fit with ordinary racial language. There is good reason to believe this if one is a realist about metaphysics, since one should not expect ordinary language to be capable of stating the best theory of the universe. In the absence of special reason to exclude race from this general claim about metaphysics, there is good reason for realists to adopt the negative thesis with respect to race. I have argued that Barnes's positive thesis is broadly correct but slightly mistaken. The metaphysics of race *is* just our best theory of race. But it is not up to metaphysicians to formulate this theory. Rather, it is up to those working in other disciplines to identify the categories of explanatory interest. This is the modified positive thesis. I will argue in Chapter 3 that there is still important work for philosophers of race to do, although it is not metaphysical.

Chapter 3: The Metaphysics, Semantics, and Ethics of Race

In Chapter 1, I argued that the best interpretation of the social constructionist claim was a constitutive one. In this chapter, I argue for *deflationism* - the view, defined in section 3.1, that the (actual) metaphysics of race debate is non-substantive, even with this constitutive interpretation of the constructionist claim. I argue in section 3.2 that the different parties broadly agree about the underlying metaphysics, and that where they do not so agree, it is not for metaphysicians of race to settle.

This doesn't imply that no debates about race are substantive. There are substantive questions concerning the metaphysics of race, as I argue in section 3.3. They concern, according to the modified positive thesis, what explains the phenomena of race. But they are to be settled by other disciplines than metaphysics. Philosophers of race do disagree about whether, and if so to what, racial terms and concepts apply, but this isn't a metaphysical dispute. Philosophers of race also disagree about the normative question of what we ought to do with race talk and thought. But this, too, isn't a metaphysical dispute, and, I argue in 3.6, isn't settled by settling the metaphysical and semantic questions.

3.1. Deflationism Defined

Deflationism, as I use the term, is a view about a given debate. It says that apparent differences of view among participants to a debate are *merely* apparent. The parties don't really disagree, or they don't disagree about the underlying, non-linguistic facts. This is to be distinguished from uses of 'deflationism' where it labels substantive first-order positions *within* a given debate, as, for instance, in philosophical debates about truth (Glanzberg 2014). Deflationism concerning the metaphysics of race is not itself a position within the metaphysics of race debate. It is a claim *about* that debate. It says that the (actual) debate is non-substantive.

Consider the claim that races exist. The racial realist (whether a biological realist or a social constructionist) says that races exist. The sceptic says that races don't exist. The deflationist says that there is something wrong with the debate. What could be wrong with the debate? Sider (2009 386-7) gives a helpful list of different versions of the deflationist thesis. The three relevant to deflationism as I have defined it are: equivocation; semantic indeterminacy;[2] and epistemic indeterminacy. *Equivocation* is the charge that the parties assert and deny different propositions using the same terms. This means that the disagreement is merely verbal. *Semantic indeterminacy* is the charge that there is no single determinate proposition expressed by 'races (don't) exist' when used by either party in the debate. Instead, the sentence is indeterminate over various candidates, some of which make the realist's claim(s) true, and some of which make the sceptic's true. *Epistemic indeterminacy* is the view that we have insufficient evidence regarding the truth-value of the sentence 'races exist'.

As Sider notes, there are versions of deflationism inspired by idealism, pragmatism, verificationism, and other versions of global anti-realism. These are not the focus of either his (2009) or this thesis. The question this chapter is exploring is whether the metaphysics of race debate is worse off as regards substantivity than other areas of metaphysics and ontology. A deflationism about metaphysics generally implies deflationism about the metaphysics of race debate. This thesis is not an attempt to vindicate realism about metaphysics in general. A certain level of realism is just assumed. Further, it is assumed that not *all* questions of ontology in particular are non-substantive, and that not *all* ontological debates are to be deflated. If one adopts a deflationism about ontology generally, this will entail deflationism about the ontology of race. This thesis is not concerned with defending ontological realism in general.

---

[2] Sider just calls this 'indeterminacy', and he calls epistemic indeterminacy 'scepticism'. I have adopted a different terminology to avoid confusion between that position and racial scepticism.

3.2. Deflationism Defended

Deflationism about the metaphysics of race is the view that the (actual) metaphysics of race debate is non-substantive. There are various ways in which the debate might be non-substantive, canvassed above. McPherson (2015), Ludwig (2015), and Mallon (2007) present arguments which support deflationism. There are significant differences between their arguments, so I will present what I take to be a sound argument in favour of deflationism, based in part on the work of these authors. If the argument is sound, either the equivocation or the semantic indeterminacy thesis is correct as regards the metaphysics of race debate.

The crux of both Ludwig's (2015) and McPherson's (2015) argument is that there are many things which 'race' and other racial terms mean or can be used to mean, on some of which 'races exist' comes out true and on others of which it comes out false, and none of these meanings is privileged as *the* meaning of 'race'. The answer to the question of whether races exist, and if they do what kind of thing they are, thus depends on which of these meanings of 'race' we adopt. The key point stressed by Mallon is that participants in the metaphysics of race debate broadly agree about which of these candidate meanings do exist. So, the disagreement is merely verbal, and the debate is non-substantive. I add that even where there is substantive disagreement between metaphysicians of race about the existence of a given candidate meaning for 'race', this is not to be settled by metaphysicians, but rather by those working in other disciplines.

What are the many things which 'race' can be taken to mean? The history of the metaphysics of race and of scientific uses of *race* demonstrates a wide variety. There are the kinds posited by scientific racialism, which Appiah (1996) identifies as *the* meaning of 'race', on which races are natural kinds with shared essences connected to normatively important traits such as intelligence and moral character. Everyone, or everyone worth listening to, agrees these don't exist. Everyone agrees that certain physical characteristics are highly correlated with continental ancestry. Beyond this, there are those who identify races with subspecies (Mayr

2002); with reproductively isolated subpopulations (Kitcher 1999, Andreasen 2000); groups linked by epidemiological propensities; groups linked by the effects of discrimination (Piper 1992); racialised groups (Haslanger 2000); groups linked by continental ancestry; groups linked by perceived continental ancestry; groups linked by genetic make-up; and so on. These don't line up or overlap significantly in their extension. Nor do they necessarily line up with the extensions of ordinary concepts. None of them can claim to be *the* referents of ordinary concepts. Some plausibly exist. Some plausibly don't. It seems implausible that one is more "joint-carving" than all of the others. So, none of them has any claim to being the dominant meaning of 'race', in virtue of being more joint-carving or in virtue of capturing the facts of ordinary usage.

The argument so far, and as presented by McPherson, is the following:

(1) There is no one dominant meaning of 'race', there are many different possible specifications.
(2) Whether 'race exists' is plausibly true depends on what 'race' is taken to mean (among these possible specifications).
(3) Therefore, the debate about whether race exists is non-substantive.

This argument is incomplete. From the claim that there is no *one* dominant meaning of 'race', it doesn't follow that there is no substantive dispute among metaphysicians when they employ the term. It could be that metaphysicians disagree about the existence of race in a specific sense of the word. For example, a biological realist might hold that races exist, when (and only when) 'race' is interpreted as applying to reproductively isolated subpopulations; a constructionist might hold that races exist when (and only when) 'race' is interpreted as applying to certain social groups; and a sceptic can hold that there is *no* sense of 'race' on which 'races exist' is true. So, another premise is required, which rules out this kind of possibility. It has to be that, once a certain specification is under consideration, the different parties to the debate wouldn't disagree about whether races, in that sense, exist, even if they

disagree about the (linguistic or conceptual) question of whether that *is* a legitimate specification of the meaning of 'race'.

The claim that once a particular specification is selected for the meaning of 'race' the different parties to the debate would no longer disagree about whether races, in that sense, exist is well supported. For example, the dispute between Kitcher (1999) and Andreasen (2000) on the one hand, and Zack (2002) and Appiah (1996) on the other, is not, as Mallon (2006) points out, about how much reproductive isolation there actually is between human subpopulations. Rather it is about "the degree of reproductive isolation required to fall under the concept "race"" (Mallon 2006 543). Mallon, perhaps rather damningly, sums up the situation:

> Skeptics says race does not exist, employing the term 'race' to mean something that everyone agrees does not exist. Constructionists insist that race does exist, again employing the term 'race' to pick out phenomena that everyone agrees exist. And [biological realists] insist that races existed and might still exist, using the term 'race' to pick out biological populations that are substantially different from the kinds whose existence [sceptics] deny. (2007 547)

A point worth noting is that by focusing on the *word* 'race', the question has moved away from individual racial identities. But shifting focus to what it is to be black, white, or Asian, is unlikely to deflect the deflationist's argument. Either the biological realists will agree with the racial sceptic that ordinary racial ascriptions are uniformly false but that biological racial ascriptions are true, or else they will hold that ordinary folk ascriptions are true ascriptions of biological properties. Social constructionists are likely to hold that ordinary racial ascriptions are true ascriptions of socially constructed properties. And sceptics will hold that ordinary racial ascriptions are all false. But the sceptic can agree that people instantiate the biological property which biological realists *take* race terms to be tracking. And biological realists can agree that what other participants to the debate *take* to be the content of a racial ascription

applies or doesn't apply, as that other participant thinks. Thus, a picture strongly supporting either the semantic indeterminacy thesis or the equivocation thesis emerges.

Unlike McPherson, Ludwig (2015) focuses on the debate between biological realists and racial sceptics. It seems clear, however, that a similar argument could be made regarding social constructionism. Ludwig follows Hirsch's (2007) general argument for ontological deflationism. The idea is that if the different parties to the debate can interpret the other side as speaking the truth in their own language by recognizing that there are different ways of talking about the subject matter at hand, the debate is non-substantive. This is a strategy familiar from pragmatists such as James (1907/1975). If racial sceptics can interpret the biological realists as saying something true in their own language, and vice versa, then the debate is non-substantive. The racial sceptic can indeed do so, because, as argued above, they could interpret the biological realists "as speaking the truth in a language in which 'race' refers to genetic clusters, patterns of mating, clades, and so on" (Ludwig 2015 260). The biological realist can interpret the sceptic as speaking the truth in a language in which 'race' refers to "subspecies, populations with visible traits that mark relevant biological differences, populations with cognitive differences, and so on" (Ludwig 2015 260).

One might worry, as Ludwig does, that this argument relies on controversial metametaphysical claims. But the argument is independent of more general metametaphysical controversies. One can accept it as applied to the metaphysics of race but not as applied, for example, to the metaphysics of composite objects. One might hold that there is a substantive disagreement there because the different sides to the debate *cannot* interpret each other as speaking the truth in each other's languages. Or one might hold that participants to that debate *are* using terms in the same way. In particular, this argument does not rely on different meanings for the existential quantifier. It relies on different meanings for 'race' and other racial terms. So, one can accept it without being a deflationist about ontology generally.

One might insist that there *is* substantive disagreement about the existence of some of the possible candidate meanings for 'race'. There are, for example, substantive disagreements about whether there are human subspecies. Mayr (2002) argues that there are, while some philosophers such as Glasgow (2019) base arguments for racial scepticism on the claim that there are not. The deflationist can reply, however, that whether there are human subspecies is a question to be settled by biological theory and empirical research. The deflationist claim should therefore be slightly more nuanced than I have so far presented it. It is not that there is no substantive disagreement between the parties to the metaphysics of race debate concerning the underlying, non-linguistic facts. It is that such disagreements are not to be settled by metaphysicians, but rather are to be settled by biologists, social scientists, and other more empirical disciplines. This is in line with the modified positive thesis advocated in the previous chapter.

This appears to be a sound argument for deflationism with respect to the metaphysics of race debate. Substantive disagreement between the parties concerning the underlying, non-linguistic, facts is limited. Where it exists, it is to be settled by researchers working in a range of other fields. The disagreement mainly concerns whether certain things fall under the concept *race* and other racial concepts. This is not a substantive metaphysical dispute, though it is an important and interesting research project for philosophers of language and race.

3.3. The Limits of Deflationism

If the negative thesis and the modified positive thesis are accepted, there is still a substantive debate, though not one for metaphysicians to participate in. The substantive debate concerns what explains certain phenomena, those associated with race. Although this is not a question for metaphysicians, it is still, in one sense, a metaphysical question, because it concerns what there is and what it is like. But this is true of science, generally. It is part of the post-Quinean paradigm that science and metaphysics are continuous. It follows, not that metaphysicians

should attempt to settle questions which properly belong in other disciplines, but that a successful scientific theory will form part of our overall successful metaphysical theory.

It is worth spelling out how one can accept deflationism with respect to the (actual) metaphysics of race debate while maintaining that there are substantive questions concerning race. Suppose one accepts that the metaphysics of race debate is non-substantive. Once a particular sense of the word 'race' is isolated, the parties to the debate will broadly agree on whether races, in that sense, exist. And no particular sense is sufficiently dominant, in virtue of being privileged by either ordinary usage or in its capacity for carving at the joints, that it can claim to be *the* meaning of 'race', thereby settling the question of whether race exists.

This claim about the lack of a dominant meaning for 'race' fits with the picture of the terrain left once the negative thesis is accepted. The negative thesis entails that the substantive question is not what the referents of ordinary race terms are. The debate need not, and probably *will* not, be conducted using ordinary race terms (with their ordinary meanings). The debate need not even be focused on what there is. Different parties may agree about what there is. Given Barnes's (unmodified) positive thesis, the debate is about what is needed to explain race. Chapter 2 argued that there is no distinctive explanatory task for the metaphysician. Instead, other disciplines will attempt to explain the phenomena of race with reference to different categories. The substantive question, then, is what is needed to explain what.

The positions within the metaphysics of race are substantive claims about what explains the phenomena of race. Biological realism becomes the claim that the phenomena associated with race admit of a biological explanation. Social constructionism becomes the claim that the phenomena associated with race admit of a social explanation. Racial scepticism becomes the claim that the phenomena of race fail to admit of systematic explanation. These positions make substantive claims on the world, some of which can be empirically tested, and some of which can be decided by criteria on theory choice which operate in these disciplines.

3.4. Pluralism, Deflationism, Scepticism

The deflationist surveys the variety of phenomena picked out by race talk, and concludes that the metaphysical debate arises out of a failure to use race talk in the same way. The pluralist surveys the same variety, and concludes that race is a multi-faceted thing. They can agree on what there is. There are versions of deflationism that are compatible with pluralism. If one is a deflationist who thinks that race exists in various senses of the word, but not others, for example, then one may be a pluralist.

Other motivations for adopting deflationism would be inconsistent with being a pluralist, however. If one thinks that race exists in exactly one sense of the word, but in other senses it does not exist, and one argues for a deflationism based on this, one can't be a pluralist. So, one can in theory be a deflationist without being a pluralist. But, in reality, deflationists are likely to think that 'races exist' is true when 'race' applies to a number of different kinds. One can also be a pluralist without being a deflationist. For example, suppose you think that races exist in the sense of 'race' employed by Haslanger (2000) and also in the sense employed by Spencer (2019a), but in no other sense. That doesn't mean that you are a deflationist. You could think that there is a substantive debate about whether there are, for example, reproductively isolated populations and about whether these are races. You just think that the answer to at least one of these is negative.

What happens to these positions when the negative thesis and the modified positive thesis are adopted? Pluralism is then the position that the correct explanation of race appeals to at least two different kinds of metaphysical structure. Maybe it appeals to more than one kind of social kind, or maybe to a social kind and a biological kind. Deflationism on the other hand, would hold that the debate (between non-metaphysicians) about the correct explanation of race is for some reason non-substantive. The contents of these two positions - pluralism and deflationism - are distinct, as are both from the content of racial scepticism. One might worry that it is hard to imagine what evidence could be brought to bear to tell in

favour of one of these views rather than another. Pluralism and scepticism both say that there is no single thing which can do all the explanatory work. So, a massively plural theory might be equivalent, evidentially, to racial scepticism. But this seems perfectly acceptable. It is just not an important question which of racial scepticism or massively plural realism would be correct in such a situation. It doesn't seem to matter what we say, if we agree on the underlying situation.

3.5. Human Kindness

The participants in the metaphysics of race debate tend to view the correct way of settling the question of whether races exist as being to decide whether there is some *kind* corresponding to our race terms. Biological realists think there are biological kinds which are races. Social constructionists think there are social kinds which are races. Sceptics think there are no kinds which are races. Racial sceptics often defend their view by arguing that, for races to exist, certain kinds would have to exist which in fact do not.

Some philosophers, such as Gannett (2010), have pointed out that this presupposes that the reality of some category consists in its being a kind. Moreover, the conception of kinds in play is often (implicitly or explicitly) the "Kripke-Putnam" conception - a collection which is united in the possession of some features, such as microphysical structure.[3] It is thought to be necessary that something possessing that feature be a member of that kind, and that any member of that kind possess that feature. The feature is often required to be intrinsic. It is also required to be relatively metaphysically simple or basic. So, there is a necessary and sufficient condition on kind-membership, which is the possession of a simple intrinsic property.

---

[3] I include the scare quotes because Kripke and Putnam actually had rather different views from each other about the metaphysics of natural kinds. To the extent that the "Kripke-Putnam" view is a view about the metaphysics of natural kinds rather than the semantics of natural kind terms, it mostly takes on Kripke's conception of them (Hacking (2007)).

With developments in the philosophy of science and the theory of natural kinds, represented, for example, in Boyd (1999), some metaphysicians of race have stopped requiring that a natural kind be such that each member possess a simple intrinsic property (Mallon 2007b). Members of a species, for example, do not share a common essence, as both Kitcher (1984) and Dupre (1981) influentially argue. They are still categories of explanatory interest to the biologist, and thus plausibly count as natural kinds. Thus, biological realists such as Kitcher (1999) and Andreasen (2000) posit biological categories characterized by extrinsic properties. Individuals are members of a race in virtue of bearing certain relations to other members of a race. Similarly, constructionist theories often posit kinds characterized by extrinsic properties, such as Haslanger's (2000), on which how one is perceived by others is a determining factor in one's racial identity. Nevertheless, the debate still turns on the question of whether there are kinds corresponding to our racial classifications. Part of the appearance of disagreement in the standard debate may be due to differing standards held by the different parties for being a kind.

Some philosophers think this approach is mistaken. Kitcher (2007) argues that his earlier (1999) argument, and many other arguments in the metaphysics of race, made incorrect presuppositions about natural kinds. Essentially, it assumed a realism about natural kinds - that they exist independently of us, our cognitive capacities, and our interests. The more recent Kitcher is a pragmatist: how we divide the world is not determined by the world. There are infinitely many acceptable divisions of the world which we are free to choose between as suits our purposes. He had earlier thought that the plurality of taxonomic practices he identifies justified a "multi-jointed" realism. He now thinks that there are so many legitimate divisions of the world that this "promiscuous realism" collapses into pragmatism. The question about whether this is promiscuous realism or pragmatism is unimportant. The important claim is that there is a panoply of ways of equally metaphysically respectable ways of dividing the world.

Applying this to the metaphysics of race debate, one could say that natural kinds are so easy to come by that some intuitively racial division of humans will be found. This isn't a theoretically interesting question precisely because natural divisions are so easy to come by. The question to focus on is the pragmatic one of whether such divisions can be drawn for legitimate purposes, and whether they should be drawn. Another way of putting the point is that there aren't any "natural" kinds, in a robust sense. No division is better than any other, as far as the world is concerned. So, we are wrong to think that we can settle anything by reference to whether certain natural kinds exist. They don't. But neither does this hand victory to the racial sceptic, on pain of embracing a global nihilism. The fact that "natural" kinds, in a robust sense, don't exist is good reason for thinking that the question is not an important one.

These views are controversial. This thesis isn't concerned with defending realism generally, or realism about natural kinds specifically, against pragmatism. Adopting the negative thesis and the modified positive thesis with respect to the metaphysics of race debate doesn't seem to affect the issues Kitcher is interested in. The question becomes which categories fulfil certain explanatory roles. If Kitcher is right, some categories will be useful for some purposes, and others for others. There will be no privileged category, independent of our contingent explanatory interests. But there will be a variety of categories which various disciplines find it useful to point to in explaining the phenomena associated with race. If Kitcher is correct, it seems to follow that the relevant questions concerning race will be what the categories of explanatory interest are, and what we ought to do with race talk. The former will still plausibly be decided by non-metaphysicians. The latter remains an interesting (non-metaphysical) question for philosophers to discuss. So Kitcher's pragmatism about kinds seems to imply similar conclusions regarding the division of the intellectual work concerning race as do the two theses.

3.6. Race Talk

Philosophers of race often discuss the metaphysics of race alongside normative issues concerning whether we ought to speak and think in terms of race. The connection is not as straightforward as philosophers of race tend to assume. They tend to think that the metaphysics settles what there is for racial terms to refer to, and, together with the correct theory of reference, thereby settles whether and to what racial terms *do* refer. This is supposed to thereby settle what we ought to do with racial terms. This is a mistaken strategy. One can hold that racial terms *do* refer but that we shouldn't use them or that they *don't* but that we should. The negative thesis implies that the terms we use to explain race, and thereby identify the metaphysically significant categories, need not be our ordinary race terms with their ordinary meanings. There can obviously be a connection between what, for example, sociology tells us about how the phenomena of race arise, particularly if our talking and thinking in terms of race plays an important part, and the normative question of what to do with race thought and talk. But the direct connection is less direct than philosophers of race often assume.

Connecting the debate about conservationism versus eliminativism with the metaphysics of race debate requires an argument, as Mallon (2006) points out. Mallon identifies a standard strategy, the *semantic strategy*, which authors often rely on to provide this connection. I present it differently, adding a step, to show two possible points of disagreement which are obscured in Mallon's presentation of the strategy:

(1) Metaphysical claim: the world has (or lacks) certain metaphysical features.
(2) Semantic claim: some particular theory of reference is correct for racial terms (or concepts).
(3) Connection between metaphysical and semantic claims: the theory of reference, together with certain other facts about racial terms, implies that these racial terms refer to those metaphysical features (or fail to refer completely).

(4) Connection to normative issues: that racial terms refer (or fail to refer) to these features means that we should (or shouldn't) use them in certain ways.

Appiah's (1996) strategy is paradigmatic. He considers two families of theories of reference. On one, terms are associated with descriptions specifying their application conditions. On another, terms (roughly) refer to what they were introduced to refer to. He argues that on neither does the world have metaphysical features which could be the referents of racial terms. So racial terms don't refer. Mallon says that someone who disagrees with Appiah on this will have to disagree with him either over what metaphysical features the world has *or* over the correct theory of reference for racial terms and concepts. Mallon's own presentation of the semantic strategy, however, obscures the possibility of denying the connection claim, that is, of denying that the theory of reference genuinely implies that racial terms apply to the metaphysical features under consideration (or denying that it implies that they fail to refer at all since their only possible referents do not exist). If Mallon is right about the broad ontological consensus among philosophers of race, which includes that Appiah is right to deny that racial essences exist, then to deny Appiah's conclusion, one must deny: (i) that the theories of reference Appiah considers exhaust the options; (ii) that these theories of reference really do imply that racial terms would have the referents Appiah claims; or (iii) that the facts about reference have the normative implications Appiah takes them to.

Social constructionists agree with Appiah's metaphysical claim that there are no biological categories the members of which have in common shared biological essences. To resist his argument for eliminativism, they must therefore dispute either (i) that the theories of reference Appiah considers exhaust the options; (ii) that these theories of reference do imply that racial terms would have the referents he claims; or (iii) that the facts about the referents of racial terms do imply that we ought not to use racial terms. Constructionists discussed above argue that there are certain social kinds, racialised groups, which Mallon plausibly includes in the ontological consensus. They argue that these are the referents of racial terms

and concepts. So, they typically dispute that the correct theory of reference does imply that the referents of racial terms are biological kinds united in the common possession of a certain racial essence. Biological realists argue that there are biological kinds and that the correct theory of reference for racial terms will have these as the referents of racial terms and concepts. The argument is thus about what something has to be like to be the referent of racial terms and concepts. The situation, then, is that participants take themselves to be having a metaphysical argument, which they hope will have normative implications settling the issue of what to do with racial terms, but often end up having an argument about the theory of reference for racial terms and concepts. This does not seem to be the most promising way of settling the issue of what to do with race talk. As Mallon argues, it leaves the answer hostage to issues in the philosophy of language. There may be no universally correct theory of reference which will settle the dispute for us. It may be underdetermined by the evidence.

The question of what we ought to do with race talk is normative. It is not clear why we should be able to go from claims about what the referents of race terms and concepts *are* to claims about how we ought to use such terms. One might hold that racial terms and concepts don't refer but that we should still use them, or that they do but that we shouldn't. To see this starkly, consider that certain terms, like racial or homophobic slurs, plausibly *do* refer to people, have a (fairly) determinate extension on a given use, but that one can recognise all of this while strongly disapproving of their use. One's reasons for saying that (at least some) people shouldn't use these terms need not be that they are semantically defective, not that they fail to refer, or that the propositional content expressed by the sentences in which they occur is false. The user of the term is not mistaken about what there is, as would someone who used 'unicorn' as if they existed. Equally, one might hold that racial terms and concepts don't refer, but that we should retain their use as a basis for solidarity and political action. Young (1990) argues for the benefits of social groups for political organization and personal identity. One might hold that only by retaining racial terms and concepts can we achieve justice, perhaps by suitably shifting their meaning so that races become *overt* social

kinds, rather than covert constructions. So, the question of what we ought to do with race talk isn't settled by identifying what our racial talk and thought refers to.

Philosophers can perform useful work examining racial language and thought. They can do both descriptive work, exploring the presuppositions and commitments of racial thought and talk, and this kind of normative work, thinking about what to do with race talk. They can also address broader normative issues concerning race. But, given the negative thesis, there is no commitment to the categories of explanatory interest lining up with the extensions of ordinary race talk. Theorists and researchers can posit categories which do not fit with ordinary race talk. And we do not have to hold that ordinary race talk does or should track categories of explanatory interest. The separation of the metaphysics of race from the normative question of what we ought to do with race talk and thought is then much more obvious. Of course, there are still likely to be connections between the metaphysics of race and the normative questions. For example, if the correct explanation of certain undesirable effects involves our use of race talk and thought, then this constitutes a good reason not to use it, or to use it differently. But this kind of connection is less direct than what is typically assumed.

Conclusion

In this thesis, I have argued that there is something wrong with the metaphysics of race debate. Even on its best interpretation, there is no substantive metaphysical dispute between its participants. Even where there is substantive disagreement between the different parties to the debate, this is to be settled by those working in disciplines other than metaphysics. The appearance of a substantive metaphysical dispute is maintained in part by the assumption that our metaphysics of race has to fit with ordinary racial language. This is not the case. Instead, our metaphysics of race is given by our best theory for explaining the various phenomena associated with race. This is for those working in other disciplines to settle, not metaphysicians. There is no work for metaphysicians of race.

To establish these claims, I argued that the best interpretation of the social constructionist claim in the metaphysics of race debate is a claim about what constitutes being a race or individual racial identity. Social constructionist metaphysicians, like other metaphysicians, aren't primarily interested in our representations of the world, but rather in the domain which those representations purport to be about. While constructionists do have commitments concerning the causal origins of racial identities, interpreting the social constructionist claim as attributing some kind of constitutive relation between human decisions, actions, practices, and institutions and races and racial identities gives the best prospects of vindicating the metaphysics of race debate.

The metaphysics of race debate is nevertheless not substantive. There *are* substantive metaphysical questions regarding race, but these aren't for the metaphysician to settle. They are still, in one sense, metaphysical questions, because metaphysics is concerned with what there is and what it is like. But the post-Quinean thought that science and metaphysics are continuous means that the questions' being metaphysical doesn't mean metaphysicians are required to answer them. Adopting the negative thesis allows us to disentangle the task of explaining race from the task of deciding what to do with race talk.

I've not established that we *ought* to adopt the negative thesis. But for those who accept the negative thesis with respect to metaphysics generally, there has to be special reason to avoid adopting it with respect to the metaphysics of race. I can see no such reason. Those who do not adopt the negative thesis but wish to retain a metaphysics of race debate for metaphysicians to engage in have to identify a substantive point of disagreement between metaphysicians of race which it is up to metaphysicians of race to settle. The prospects for this seem doubtful. It would be better for philosophers to focus on conceptual clarification, ordinary language analysis, and the normative questions around race talk and other social, moral, and political issues. Many metaphysicians of race are already explicitly or implicitly focusing on such questions. More clarity might be achieved by abjuring the label 'metaphysics'.

Bibliography

Andreasen, R. 2000. "Race: Biological Reality or Social Construct?" *Philosophy of Science* 67: 653-666.

Andreasen, R. 2004. "The Cladistic Race Concept: A Defense" *Biology and Philosophy* 19:425-442.

Andreasen, R. 2005. "The Meaning of 'Race': Folk Conceptions and the New Biology of Race" *The Journal of Philosophy* 102(2): 94-106.

Anscombe, G. E. M. 1957. Oxford: Blackwell.

Appiah, K. A. 1996. "Race, Culture, Identity, Misunderstood Connection" in K. A. Appiah and A. Gutmann (eds.) *Color Conscious: The Political Morality of Race*. Princeton, NJ: Princeton University Press. 30-105.

Asta. 2018. *Categories We Live By*. Oxford: Oxford University Press.

Barnes, E. 2016. *The Minority Body*. Oxford: Oxford University Press.

Barnes, E. 2020. "Gender and Gender Terms" *Noûs* 54(3): 704-730.

Bennett, K. 2017. *Making Things Up*. Oxford: Oxford University Press.

Boyd, R. 1999. "Homeostasis, species, and higher taxa" in R. Wilson (ed.) *Species: New Interdisciplinary Essays*. Cambridge, MA: MIT Press. 141-86.

Burchard et al. 2003. "The Importance of Race and Ethnic Background in Biomedical Research and Clinical Practice" *New England Journal of Medicine* 348(12): 1170-5.

Butler, J. 1990. *Gender Trouble: Feminism and Subversion of Identity*. New York: Routledge.

Cavalli-Sforza, L. and Feldman, M. 2003. "The Application of Molecular Genetic Approaches to the Study of Human Evolution" *Nature: Genetics* 33: 266-75.

Cooper and David. 1986. "The Biological Concept of Race and its Application to Public Health and Epidemiology" *Journal of Health Politics, Policy, and Law* 11(1): 97-116.

DeRose, K. 1995. "Solving the Skeptical Problem" *The Philosophical Review* 104(1): 1-52.

Dodd, J. 2007. *Works of Music: An Essay in Ontology*. Oxford: Oxford University Press.

Duhem, P. 1906/1954. *The Aim and Structure of a Physical Theory*. Princeton: Princeton University Press.

Dupre, J. 1981. "Natural Kinds and Biological Taxa" *The Philosophical Review* 90(1): 66-90.

Economist, The. 2020, 11 June "Free Exchange: Stony the Road" https://www.economist.com/finance-and-economics/2020/06/11/economic-research-documents-black-americans-struggle-for-equality [accessed 13 June 2020.]

Edgington, D. 1995. "On Conditionals" *Mind* 104: 235-329.

Foucault, M. 1972. *The Archaeology of Knowledge*. New York: Pantheon Books.

Foucault, M. 1978. *The History of Sexuality, Vol. I, An Introduction*. New York: Pantheon Books.

Fricker, M. 2007. *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford: Oxford University Press.

Gannett, L. 2010. "Questions Asked and Unasked: How by worrying less about the 'really real' philosophers of science might better contribute to debates about genetics and race" *Synthese* 177: 363-85.

Garrod, J. 2006. "A Brave Old World: An Analysis of Scientific Racism and BiDil" *McGill Journal of Medicine* 9(1): 54-60.

Glanzberg, M. 2014. "Truth" in Edward Zalta (ed.) *Stanford Encyclopedia of Philosophy*.

Glasgow, J. 2009. *A Theory of Race*. New York: Routledge.

Glasgow, J. 2019. "Is Race an Illusion or a (Very) Basic Reality?" in Glasgow et al. *What is Race? Four Philosophical Views*. Oxford: Oxford University Press. 111-49.

Hacker, A. 1992. *Two Nations: Black and White, Separate, Hostile, Unequal*. New York: Schribner.

Hacking, I. 1995. "The Looping Effects of Human Kinds" in D. Sperber, D. Premack, and A. J. Premack (eds.) *Causal Cognition: A Multidisciplinary Debate*. New York: Clarendon Press. 351–94.

Hacking, I. 1999. *The Social Construction of What?* Cambridge, MA: Harvard University Press.

Hacking, I. 2007. "Putnam's Theory of Natural Kinds is not the Same as Kripke's" *Principia* 11(1): 1-24.

Hall, S. 1996. "New Ethnicities" in D. Morley and K.-H. Chen (eds.) *Stuart Hall: Critical Dialogues in Cultural Studies*. London: Routledge. 442-51.

Haslanger, S. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34: 31-55.

Haslanger, S. 2012. *Resisting Reality: Social Construction and Social Critique.* Oxford: Oxford University Press.

Haslanger, S. 2012a. "Social Constructionism: The "Debunking" Project" in *Resisting Reality*. Oxford: Oxford University Press. 113-38.

Haslanger, S. 2012b. "Ontology and Social Construction" in *Resisting Reality*. Oxford: Oxford University Press. 83-112.

Haslanger, S. 2019. "Tracing the Sociopolitical Reality of Race" in Glasgow et al. *What is Race? Four Philosophical Views*. Oxford: Oxford University Press. 5-37.

Hawthorne, J. 2004. *Knowledge and Lotteries*. Oxford: Oxford University Press.

Hayward, C. 2000. *De-Facing Power*. Cambridge: Cambridge University Press.

Hirsch, E. 2007. "Ontological Arguments: Interpretive Charity and Quantifier Variance" in T. Sider, J. Hawthorne, and D. Zimmerman (eds.) *Contemporary Debates in Metaphysics*. Oxford: Blackwell. 367-91.

James, W. 1907/1975 *Pragmatism: A New Name for Some Old Ways of Thinking*. Cambridge, MA: Harvard University Press.

Jenkins, K. 2016. "Amelioration and Inclusion: Gender Identity and the Concept of Woman" *Ethics* 126(2): 394-421.

Johnston, D. and Lordan, G. 2015. "Racial Prejudice and Labour Market Penalties during Economic Downturns" *European Economic Review* 84: 57-75

Kitcher, P. 1984. "Species" *Philosophy of Science* 51(2): 308-333.

Kitcher, P. 1999. "Race, Ethnicity, Biology, Culture" in L. Harris (ed.) *Racism*. New York: Humanity Books. 87-120.

Kitcher, P. 2007. "Does 'Race' Have a Future?" *Philosophy and Public Affairs* 35(4): 293-317.

Kripke, S. 1980. *Naming and Necessity*. Cambridge, MA: Harvard University Press.

Lewis, D. 1969. *Convention: A Philosophical Study*. New Castle: Basil Blackwell.

Ludwig, D. 2015. "Against the New Metaphysics of Race" *Philosophy of Science* 82(2): 244-65.

Machery, E. 2014. "Social Ontology and the Objection from Reification" in M. Gallotti and J. Michael (eds.) *Perspectives on Social Ontology and Social Cognition*. Springer. 87-102.

MacKinnon, C. (1989) *Toward a Feminist Theory of the State*. Cambridge, MA: Harvard University Press.

Mallon, R. 2004. "Passing, Travelling, and Reality: Social Constructionism and the Metaphysics of Race" *Noûs* 38(4): 644-673.

Mallon, R. 2006. "'Race': Normative, not Metaphysical or Semantic" *Ethics* 116(3): 525–51

Mallon, R. 2007a. "A Field Guide to Social Construction" *Philosophy Compass* 2(1): 93-108.

Mallon, R. 2007b. "Human Categories Beyond Non-Essentialism" *Journal of Philosophy* 15(2): 146-68.

Mallon, R. 2016. *The Construction of Human Kinds*. Oxford: Oxford University Press.

Mallon, R. 2017. "Social Roles and Reification" in J. Kiverstein (ed.) *Routledge Handbook of Social Mind*. London: Routledge. 431-446.

Mallon, R. 2019. "Naturalistic Approaches to Social Construction" in Edward Zalta (ed.) *Stanford Encyclopedia of Philosophy*

Mayr, E. 2002. "The Biology of Race and the Concept of Equality" *Daedalus* 131: 89-94.

McPherson, L. K. 2015. "Deflating "Race"" *Journal of the American Philosophical Association* 1(4): 674-93.

Mills, C. 1998. *Blackness Visible: Essays on Philosophy and Race*. Ithaca: Cornell University Press.

Omi, M. and Winant, H. 2014. *Racial Formation in the United States*. New York: Routledge.

Office for National Statistics. 2012. "Ethnicity and National Identity in England and Wales: 2011" https://www.ons.gov.uk/peoplepopulationandcommunity/culturalidentity/ethnicity/articles/ethnicityandnationalidentityinenglandandwales/2012-12-11 [accessed 3 June 2020.]

Piper, A. 1992. "Passing for White, Passing for Black" *Transition* 58: 4-32.

Putnam, H. 1975. "The Meaning of "Meaning"" in *Mind, Language, and Reality: Philosophical Papers*. New York: Cambridge University Press. 215-71.

Root, M. 2000. "How We Divide the World" *Philosophy of Science* 67: 628-639.

Searle, J. 1995. *The Construction of Social Reality*. New York: The Free Press.

Shulman, J. and Glasgow, J. 2010. "Is Race Thinking Biological or Social, and Does it Matter for Racism? An Exploratory Study" *Journal of Social Philosophy* 41(3): 244-59.

Sider, T. 2009. "Ontological Realism" in D.J. Chalmers, D. Manley, and R. Wasserman (eds.) *Metametaphysics*. Oxford: Oxford University Press.

Sider, T. 2011. *Writing the Book of the World*. Oxford: Clarendon Press.

Spelman, E. 1988. *Inessential Woman: Problems of Exclusion in Feminist Thought*. Boston: Beacon Press.

Spencer, Q. 2019a. "How to Be a Biological Racial Realist" in Glasgow et al. *What is Race? Four Philosophical Views*. Oxford: Oxford University Press. 73-110.

Spencer, Q. 2019b. "A More Radical Solution to the Race Problem" *Aristotelian Society Supplementary Volume* 93(1): 25-48.

Stanley, J. 2005. *Knowledge and Practical Interests*. Oxford: Oxford University Press.

Wagner et al. 2017. "Anthropologists' Views on Race, Ancestry, and Genetics" *American Journal of Physical Anthropology* 162(2): 318-327.

Williamson, T. 2000. *Knowledge and its Limits*. Oxford: Oxford University Press.

Williamson, T. 2007. *The Philosophy of Philosophy*. Oxford: Blackwell.

Young, I. M. 1990 *Justice and the Politics of Difference*. Princeton: Princeton University Press.

Zack, N. 1993. *Race and Mixed Race*. Philadelphia: Temple University Press.

Zack, N. 2002. *Philosophy of Science and Race*. New York: Routledge.