

## Arguments from Reference and the Worry About Dependence<sup>1</sup>

"Arguments from reference" are arguments that employ the assertion of a substantive theory of reference as a premise in an argument with a philosophically significant conclusion.<sup>2</sup> Recent work has challenged the viability of these arguments on the basis of empirical evidence of variation in the intuitions that substantive theories of reference are usually taken to be responsible to (Mallon et al. ms, Machery et al. 2004). My present aim is to raise a distinct (though complementary) sort of concern with the use of theories of reference in philosophical discourse and then to consider the possibility of empirically validating this concern by reference to a novel sort of "quantitative" empirical approach suggested recently by Shaun Nichols (forthcoming).

The concern I raise is whether the particular *theories of reference* or *reference relations* (I use these terms interchangeably in what follows) that are employed in particular philosophical discussions are actually chosen with a view to entailing or accommodating a desired philosophical outcome. I argue below that such *dependent* selections of assumptions about reference give us little reason to think the assumptions are true. I go on to argue that if we became convinced that such assumptions really are chosen simply to ensure a desired outcome, it would give us reason for skepticism about arguments from reference since it would undermine our sense that such arguments tracked any independent truth about the reference of our words or concepts.

---

<sup>1</sup> Many thanks to Jeff Buechner, Rafaella DeRosa, Anna Stubblefield, and Shaun Nichols, for comments on earlier versions of this work, as well as to Edouard Machery and Stephen Stich for fruitful collaborations on earlier papers out of which this one grew. Special thanks to members of Jonathan Weinberg and members of his Experimental Epistemology Lab (including Josh Alexander, Chad Gonnerman, and Kari Theurer) for perceptive criticisms of an earlier draft.

<sup>2</sup> Following Mallon et al. (ms), I use "substantive" here to exclude deflationary accounts of reference (e.g. Field 1986, 1994; Horwich 1994).

10/1/07

But how can we tell the reasons why assumptions about reference are selected? The obvious way is simply to look at what individual authors say. Do they mean to simply accommodate their own desired philosophical views? Or do they present entailment from a selected theory of reference (along with other premises) as providing an independent reason to believe their conclusions are true? This interpretive method, however, faces two difficulties. First, while the employment of theories of reference in philosophical discussion is common, individual texts are often unclear about what argumentative role they are intended to play. Second, even where individual authors explicitly intend the invocation of a substantive theory of reference to provide independent support for a philosophically significant conclusion (i.e. in arguments from reference), we might wonder whether the substantive theory really enjoys independent justification or whether its details are selected primarily with a view to ensuring a desired philosophical view.

It is this second sort of difficulty I consider further below. I suggest that straightforward interpretation, focusing on individual authors, simply cannot tell us whether or assumptions about reference have been systematically chosen on grounds that do not justify them.<sup>3</sup> Instead, I consider the possibility for exploring this question via Nichols's "quantitative approach" to philosophy. Nichols writes: "The guiding idea is that a more abstract, quantitative approach can reveal patterns that get lost in the traditional project of close readings. In much empirical research, one tries to detect patterns through the noise of individual variation" (forthcoming, ms 3-4). Here what we want to know is whether assumptions about reference in philosophical discussions are *typically* selected independently of the desirability of the conclusions to be drawn. While

---

<sup>3</sup> To say they were chosen for reasons that do not justify them is not, by itself, to say there are no reasons to justify them. We revisit this issue below.

10/1/07

the attempts below to apply a quantitative approach to this question reveal the difficulty of applying this sort of reasoning to a contemporary philosophical discussion, they provide some support for doubts about the selection of reference relations.

Here is how I proceed. In section 1, I discuss what arguments from reference are, and how they are used to achieve substantive philosophical conclusions. Then, in section 2, I discuss the possibility that theories of reference may be chosen simply to ensure a desired philosophical outcome, and I discuss why that calls their usefulness into question. In section 3, I look for quantitative evidence for the independence or dependence of assumptions about reference, finding some in a case study of the arguments from reference used in the metaphysics of race. Finally, in section 4, I discuss how to go forward if these suspicions about the dependence of assumptions about reference are correct.

## **1. Arguments from Reference**

### **1.1. How Reference Matters**

Substantive theories of reference have currency far outside the philosophy of language, figuring prominently in discussions ranging from the philosophy of science to ethics.<sup>4</sup> Theories of reference are employed in these diverse areas because they can be used to move from claims about particular terms to philosophically significant claims about the world.

---

<sup>4</sup> See Mallon et al. for more careful discussion of arguments from reference in the philosophy of mind, the philosophy of science, the philosophy of race, and meta-ethics.

10/1/07

Mallon et al. (ms) identify three steps in an argument from reference.<sup>5</sup> First, a substantive theory of reference is assumed for a term *t* (or a class of terms *T*). Crucially, this theory of reference is taken to be true, on some independent grounds. Second, a claim about the reference of the term *t* (or class of terms *T*) is made. For example, it might be claimed that *t* refers or fails to refer. Or a particular referent of *t* might be identified (i.e. *t* refers to *t*). Finally, a philosophically significant conclusion is drawn, for example, that the referent of *t* exists, or fails to exist.

By way of illustration, consider one much discussed example from the philosophy of mind with this structure.<sup>6</sup> Eliminativists in the philosophy of mind (Churchland 1981, Stich 1983) argued that propositional attitudes like beliefs and desires were posits of a folk psychological theory used for predicting and explaining behavior. They went on to argue that this folk psychological theory was massively mistaken, as revealed by the emerging scientific study of the human mind and brain. So they concluded that beliefs and desires do not exist.

Such an argument implicitly assumes that the terms "belief" and "desire" are defined by their theoretical role in folk psychology. So if this theory is massively mistaken, then these terms will not refer, and their referents will not exist. In other words, the argument implicitly assumes some version of a descriptivist theory of reference (e.g. Lewis 1970, 1972), whereby a term refers only if most of the description users associate with the term turns out to be true.<sup>7</sup> Schematically, descriptivist theories of reference hold that:

---

<sup>5</sup> Cf. Bishop and Stich's (1998) discussion of the "flight to reference."

<sup>6</sup> This example is discussed in Stich 1996, Bishop and Stich 1998, Mallon et al. ms.

<sup>7</sup> So while Churchland, for example, does not frame his argument in terms of a theory of reference, we charitably read him as assuming one and as producing an argument from reference.

10/1/07

D1. Competent speakers associate a description with a term *t*. This description specifies a set of properties.

D2. An object is the referent of *t* if and only if it uniquely or best satisfies the description associated with it.<sup>8</sup>

Famously, however, this account of reference has competitors: most notably, causal-historical theories of reference of the sort defended by Kripke (1972) and Putnam (1975).

While such theories may take a variety of forms, schematically, they hold that:

C1. A term *t* is introduced into a linguistic community for the purpose of referring to a particular thing (e.g. a person or a property). The term continues to refer to that thing as long as its uses are linked to the thing via an appropriate causal chain of successive users: every user of the term acquired it from another user, who acquired it in turn from someone else, and so on, back to the first user who introduced the term.

C2. Speakers may associate descriptions with terms. But after the term is introduced, the associated description does not play any role in the fixation of the referent. The referent may entirely fail to satisfy the description.

The fact that descriptivist theories of reference have competitors is important, because it means that one can resist the philosophically significant conclusion of an argument from reference by asserting an alternative theory of reference. This point was made in response to the eliminativist argument about propositional attitudes by William Lycan:

I am entirely willing to give up fairly large chunks of our commonsensical or platitudinous theory of belief or desire (or of almost anything else) and decide that we

---

<sup>8</sup> This, and the schematic statement of a causal historical theory stated below, are taken from Mallon et al. ms.

10/1/07

were just wrong about a lot of things, without drawing the inference that we are no longer talking about belief or desire. To put the matter crudely, I incline away from Lewis's Carnapian and/or Rylean cluster theory of reference of theoretical terms, and toward Putnam's (1975) causal-historical theory. (Lycan 1988, 31-32)

By asserting a causal-historical theory of reference in place of a descriptivist one, Lycan was able to resist the eliminativist conclusion about beliefs and desires. He simply substituted one argument from reference for another.<sup>9</sup>

Other examples of arguments from reference are not hard to find.<sup>10</sup> For example, in the philosophy of race (an area we discuss further below), racial skeptics (Appiah 1995, 1996; Zack 1993; Blum 2002) have argued that race does not exist on the grounds that nothing satisfies the central elements of the folk theory of race. Like eliminativists in the philosophy of mind, these skeptics about race employ something like a descriptivist theory of reference in the attempt to establish that race does not exist. But, here again, one can resist the conclusion by resisting the theory of reference. Robin Andreasen (2000), who defends an account on which a race is a natural population, does this by pointing out the alternative of a causal-historical interpretation of the relevant terms, writing that "The objectivity of a kind, biological or otherwise, is not called into question by the fact that ordinary people have mistaken beliefs about the nature of that kind.

Those familiar with the causal theory of reference for natural kind terms will be aware of

---

<sup>9</sup> This discussion echoes Stephen Stich's (1996), in which he takes Lycan's observation as the basis for a thorough rethinking of his own eliminativism.

<sup>10</sup> E.g. Mallon et al. (ms) discuss further examples from the philosophy of science, ethics, and the philosophy of race.

10/1/07

this possibility" (S62).<sup>11</sup> Here, as in the case of the propositional attitudes, different philosophically significant conclusions follow from different assumptions about reference.

## 1.2. How Auxiliary Assumptions Also Matter

Of course, the two schematic approaches to reference we have considered so far (descriptive, causal-historical) are only rough outlines, and there are many ways of developing these theories, as well as combinations of – or alternatives to – these theories. Even when we focus just on these two schematic approaches, there are a great many additional assumptions necessary in order to use the theories of reference to determine the referents of particular terms. And just as our choice of a general schematic approach to reference can change the philosophically significant outcome of an argument from reference, so the different choices of *auxiliary assumptions* within these theoretical approaches can also have philosophically significant results.

Consider the simple descriptivist theory we described above. To apply this theory to a real case, one must decide exactly what the description is that fixes the referent of a term and how much of which components of the description a thing would have to satisfy in order to count as the referent. And to employ causal-historical theories one must decide on just what thing bears the proper causal-historical relationship to current uses of the term and whether or not the reference has switched to a new thing somewhere along the way (e.g. Evans 1973, Tye 1998). Some auxiliary assumptions must be made by both

---

<sup>11</sup> Though here and later (Andreasen 2005) Andreasen insists that her point does not depend on the truth of the causal historical theory. Other accounts of reference could accommodate the same point. See 2.2. below.

10/1/07

sorts of theories, including assumptions addressing the role (if any) that deference to experts or to the community should play in fixing a referent (Putnam 1975, Burge 1979).

The important point, for present purposes, is that just as the choice of a general theory of reference can alter what philosophically significant outcome follows, so different auxiliary assumptions can also make a big difference to just what conclusion can be drawn from an argument from reference – a fact well illustrated by the use of theories of reference in recent discussions of the metaphysics of race.

Above, we noted that racial skeptics have argued that race does not exist, while racial naturalists, like Andreasen, believe that race does exist as a biological kind. A third metaphysical alternative has been offered by *racial constructionists* (e.g. Mills 1998, Taylor 2000, 2004) who hold that race is real but not biological. Rather, it is some sort of socially or culturally produced kind. Proponents of each of these three positions have produced arguments from reference to defend their accounts. Crucially, however, the outcome of these various arguments depends closely not only on the broad schematic assumptions made about reference (e.g. descriptivist or causal historical), but also on the particular auxiliary assumptions made.

Because we will return to considering the use of arguments from reference in the metaphysics of race below, it will be worthwhile to consider a few of these arguments in a bit of detail. Consider first the employment of descriptivist theories of reference. As we noted above, skeptics about the existence of race have often cast their arguments in terms of descriptivist accounts of reference. They typically insist that the ordinary concept of race entails some or another features that turn out not to be true of any entity, so the term does not refer, and race does not exist. So, for example, K. Anthony Appiah,

10/1/07

Naomi Zack, and Lawrence Blum have all offered analyses of the concept of race on which it entails a false commitment to the view that races have significant, intrinsic, natural differences. Because this folk account of race is seriously in error, they conclude that "race" does not refer, and race does not exist.

We already noted that one could resist this argument by resisting the descriptivist theory on which the analysis depends, but one can also resist this argument simply by choosing alternative auxiliary assumptions. For example, Philip Kitcher (1999) as well as Massimo Pigliucci and Jonathan Kaplan (2003) defend naturalist accounts of race by analyzing the way scientists use the concept, resisting eliminativism by employing an alternative description. This strategy also has been employed in defense of constructionism by Paul Taylor (2000) who suggests employing a description (suggested in work by W.E.B. Du Bois) that he thinks lacks the troublesome implications of the descriptions skeptics suggest. An alternative way of resisting eliminativism is to select a relatively *austere* description, one on which the description contains fewer elements.<sup>12</sup> Charles Mills (1998) and Michael Hardimon (2003) choose this tactic, suggesting that the ordinary concept of race can be divorced from elements of its racist and false conception by choosing a description that is simply "thinner," containing fewer elements than those put forward by racial skeptics (what Hardimon calls the "logical core" of the ordinary concept). Because it contains only a subset of the elements skeptics associate with "race" and does not contain the false elements that skeptics include, this thinner description is easier to satisfy.<sup>13</sup> Together, these cases illustrate that accepting a broadly descriptivist

---

<sup>12</sup> This use of "austere" comes from Mallon and Stich (2000) though the term is adapted from Horgan and Graham (1990).

<sup>13</sup> For more on this interpretation of Mills, see Mallon 2004, 2006.

10/1/07

approach to theories of reference leaves substantial auxiliary assumptions unspecified, and differences in these auxiliary assumptions make a difference in what philosophically significant conclusions can be drawn.

A causal-historical approach offers similar flexibility. Here again, racial skeptics have insisted that applying a causal historical theory of reference to racial terms will show that the terms do not refer, and so race does not exist. But constructionists have suggested that a causal historical approach might allow that racial terms refer to social constructions, despite the fact that this is contrary to common sense (Haslanger 2003, 2005). And as we saw above, Andreasen has defended a naturalist account of race by appealing to a causal theory of reference. These cases show that auxiliary assumptions can also significantly influence what conclusions can be drawn from a causal historical theory.<sup>14</sup>

In short, arguments from reference are a common means of moving from claims about theories of reference to philosophically significant conclusions. But they can be disputed, both by disputing the general accounts of reference they employ, but also by contesting the particular auxiliary assumptions required to apply these general accounts to actual discourse.

## **2. Independent and Dependent Reasons**

### **2.1. Choices of a Theory of Reference**

Arguments from reference ostensibly involve the *independent* selection of a theory of reference and attendant assumptions that are subsequently brought to bear on the domain

---

<sup>14</sup> Indeed, the dispute between Glasgow (2003) and Andreasen (2005) can be seen as a dispute over the right auxiliary assumptions.

10/1/07

in question. Call the combination of a theory of reference with its attendant auxiliary assumptions a *complete* theory of reference. The selection of a complete theory of reference R for a term *t* (or class of terms T) is *independent* for a person *p* if *p*'s reason or reasons for choosing R are independent of the desirability of R's being true. This way of speaking cashes out the independence of the selection of a theory in terms of the independence of the reason for choosing it. While giving a full account of this latter sort of independence is difficult and would take us too far a field, I discuss the purpose of distinguishing such reasons in a moment.

First, let's consider one possible independent basis for choosing a complete theory of reference: its fit with the folk's referential intuitions. In the philosophy of language, accounts of reference are typically justified by reference to intuitions about how terms refer in actual and possible cases. One way of interpreting this project of producing a complete theory of reference is along the lines of the Chomskian project in linguistics: intuitions about reference provide evidence of an implicit folk theory of reference. On the basis of philosophical practice in reconstructing this implicit theory, we can guess that it is usually assumed to be universal, at least among speakers of the same language.<sup>15</sup> On this *quasi-Chomskian* interpretation, we use the intuitions as evidence in the attempt to reconstruct this implicit theory, and the best complete theory of reference will be the one that best accounts for these intuitions. If this general picture is correct – including the substantial empirical assumption that people have an implicit theory of reference that generates their intuitions about actual and possible cases (Stich 1996, 85-6, endnote 35) –

---

<sup>15</sup> But see Machery et al. 2004.

10/1/07

then it offers us an account of what an independent basis for choosing a complete reference relation could be: the intuitions generated by our implicit folk theory.<sup>16</sup>

In contrast, the choice of a complete theory of reference or reference relation *R* for a term *t* (or class of terms *T*) will be *dependent* for a person *p* if *p*'s reason for choosing *R* is that *R*'s being true is desirable. For example, in choosing a reference relation for moral terms like “good,” “right,” or “just,” we might motivate our choice by appeal to the desirability of using that relation in the moral domain. We might, for example, follow Michael Smith in employing a theory of reference (at least in part) because it allows us to preserve and explain our sense that there are *a priori* moral truths. He writes that,

There is in fact a rich set of platitudes about rightness that those who want simply to fix the reference of rightness by some minimal reference-fixing description simply fail to take into account. [*E.g.* Causal theories of reference] are therefore unable to accommodate or explain these *a priori truths*” (Smith 1994, 32).

Taken in isolation, this consideration would make a person's choice of the reference relation for moral terms dependent on the philosophical consequences for morality. Specifically, that person might adopt a certain variety of descriptivist account because the account entails the desired conclusion that there is a rich set of necessary and *a priori* moral truths.

Dependent selections of theories of reference or auxiliary assumptions look problematic because the sort of reasons they rely on (*e.g. it would be good if R were*

---

<sup>16</sup> While the most common way of independently motivating a theory of reference in the philosophy of language is to appeal to intuitions about the reference of a class of terms, there could be other independent bases for selecting a theory of reference, a point we return to below.

10/1/07

*true*), do not seem to be the right kinds of reason to justify the theory or assumption.<sup>17</sup>

While it is difficult to distinguish the wrong kinds of reason from the right ones in a precise way, the general idea is clear: a proposition's desirability is generally independent of its truth. So the fact that it would be desirable for a proposition to be true is the wrong kind of reason to endorse it. In our case, just because a reference relation or assumption has a desirable entailment does not make it true, and it gives us no reason to think that it is true.

## **2.2. Interpreting Arguments from Reference**

While we can conceptually distinguish the independent selection of a theory of reference from the dependent selection, distinguishing these types in philosophical practice is often more difficult. Because authors themselves often introduce theories of reference, implicitly or explicitly, without explaining exactly why they have chosen that theory of reference, a question may arise as to whether they have chosen the reference relation for independent reasons or because of its beneficial results for the domain in question. Consider, for example, Richard Boyd's (1988) famous articulation of a strategy for defending moral realism. A central part of Boyd's approach is the application of a causal-historical account of reference to moral terms, and the assertion that such an application would solve various problems for moral realism in a way similar to the way

---

<sup>17</sup> The phrase the "wrong kind of reason" is found mostly in discussions of reasons for values (e.g. Olson 2004). Here, I follow Pamela Hieronymi (2005) in using the phrase to contrast with proper reasons for *beliefs*. As she observes, "there is a quite general problem about identifying the appropriate reasons for attitudes, a problem that is not restricted to reasons for those attitudes involved in valuing" (2005, 438, fn. 2).

10/1/07

causal-historical theories addressed various problems for scientific realism.<sup>18</sup> It is far from clear, however, the exact extent to which Boyd thinks that the causal historical theory is itself supposed to support realism about moral terms, and the extent to which the goal of preserving and defending moral realism is supposed to motivate the adoption of a causal-historical approach to the reference of terms.

Boyd is not alone in not making his reason for selecting a theory of reference for moral terms explicit, but the most plausible way to understand the uses of theories of reference in most philosophical contexts is that such invocations are intended to recommend well-known theories for which there are independent reasons. It is because these (often unspecified) reasons make it the case that the theory of reference is true, that the theory itself becomes a valuable premise in an argument from reference.<sup>19</sup>

We can see this commitment when we realize that arguments from reference usually employ some well known approach (typically one corresponding to one of our schema from section 1.1.). If philosophers only wanted to show that a proposition was compatible with *some* theory of reference, then it would be very easy to specify an arbitrary theory of reference to do the job. For example, it would be very easy to select a complete theory of reference on which propositional attitudes exist or do not exist. As long as there exists *some possible* reference relation that could serve the desired theoretical role, then that relation could be selected using its success in the role as the

---

<sup>18</sup> For example, Boyd thinks that the *a posteriori* definition of moral terms provided by causal-historical theories of reference provides a way of identifying goodness with a natural property while avoiding the "naturalistic fallacy" (1988, 199) - it would therefore provide a means for a naturalist to answer G.E. Moore's open question argument.

<sup>19</sup> In an early employment of an argument from reference, Robert Merrihew Adams (1979) draws this sort of connection, suggesting that the independent success of the Kripke/Putnam view of natural kind terms "will enhance the plausibility of an analogous treatment of the nature of right and wrong" (1979, 73).

10/1/07

primary criterion. The universe of possible reference relations is enormous. Cherry picking a relation to suit some need is simply not difficult. In practice, philosophers do not typically choose bizarre or contrived theories of reference in arguing for philosophically significant conclusion. Rather, they aim for the theory they choose to be independently justified.

Even in cases where a philosopher stops short of recommending a single theory of reference, something like this is going on. Returning to our examples from the metaphysics of race, while Andreasen (2000) argues that causal-historical accounts of reference could allow race to be natural, she resists tying her argument to the details of that particular theory of reference. And Appiah (1996) argues that race skepticism follows from *both* descriptivist and causal-historical theories of reference, because he does not think either theory entirely captures the true theory of reference. While these theorists stop short of a straightforward endorsement of the theories of reference they invoke, their practice indicates that they take some theories to be more justifiable than others, and they present their conclusions as following from these most plausible accounts. The assumption behind these arguments is that these familiar theories encompass the plausible candidates for a correct theory of reference (perhaps because they cohere with clusters of strongly held intuitions). Just as some theorists hold a single theory to be independently justifiable, these theorists hold a family of theories (or even several families) to be independently justifiable.

### **2.3. An Author's Reasons and Independent Justifications**

Whatever their own reasons for (dependently or independently) selecting a theory of reference, philosophers are interested in selecting a theory or theories of reference for

10/1/07

which independent reasons can be given. But to say that a philosopher employs a theory of reference for which independent reasons may be given is not to say the philosopher selected that theory because of those reasons. Rather, it could be the case that philosophers usually choose a theory of reference and its assumptions dependently (e.g. because it entails a desired conclusion), but choose from general approaches for which they think some independent justification can be given.

But this strategy can only take one so far. While it is more than plausible to think that independent reasons exist for employing one or another of the referential schema we discussed above (because, among other things, there are well known arguments for each approach and extensive literatures populated by famous philosophers advocating the merits of each that could be consulted), it is far less obvious what independent reasons can be given for making the various auxiliary assumptions in one way rather than another. There simply is not the same consensus on the right way or ways to answer these questions in particular cases. So for example, in attempting to use an argument from reference to decide on the correct referent for "race," we find a public discourse in philosophy full of independent reasons for assuming a descriptivist or a causal historical theory of reference. But in sharp contrast, we find no such public library of independent reasons (i.e. reasons that are independent of the desirability of the conclusions that could be drawn from the assumption) on whether, say, a racial description necessarily involves a commitment to intrinsic biological difference. So, even while we recognize that there exists a public discourse full of independent reasons for some very general assumptions about reference relations, we have to ask whether there are any independent reasons to think that the many more specific assumptions that arguments from reference make about

10/1/07

theories of reference are correct. This turns out to be important, since (as we saw in section 1.2) these assumptions can affect the conclusion of an argument from reference. If they turn out to lack justification, so will the conclusions of arguments from reference.

Three possibilities for how the assumptions about reference might be supported will concern us. First is the possibility we just mentioned: philosophers choose a complete theory of reference because it is desirable (e.g. it has desirable entailments), but they choose a theory from among those for which independent reasons can be readily given. I just suggested this first possibility was plausible in thinking about what schematic approach to adopt, but far less plausible in thinking about the various assumptions about auxiliary assumptions to be made. This leads us to the second, worrisome possibility that philosophers are making these assumptions dependently (to ensure the outcome they want) and no independent grounds justify the choice. If this turns out to be the case, then we would have reason to ignore arguments from reference as ornaments that decorate the conclusions they seem to prove.

There is a third alternative, however. It could be that philosophers choose the various assumptions of a complete theory on independent grounds even though these are not always clear. One possibility is that philosophers make these assumptions on the grounds of their own independent intuitions about the reference of the term or terms in question. For example, Michael Hardimon writes, "The ordinary concept of race is *our* concept. As competent users of the term, we are *entitled* to draw on our intuitions about its application" (2003, 441). Hardimon's intuitions could be an independent basis for his assumptions about the reference of the ordinary concept of race, even if the independence of these intuitional judgments is not obvious. If this third possibility were the case, then

10/1/07

perhaps choices about reference are justified, or could be justified by making these grounds more clear. In the remainder of this section, and throughout the next, I will give reasons to doubt that such auxiliary assumptions about reference generally have such independent grounds.

#### **2.4. Worrying About Dependence**

If the auxiliary assumptions made by philosophers tended to converge on one or a few central ways of employing theories of reference, that would give us some reason to think that these assumptions were tracking some independent fact (even if it was not obvious what the fact is), but instead (and as we saw above in our examination of the debate over the metaphysics of race) there is considerable variation in the sorts of assumptions that individual theorists make.

This variation is somewhat troubling if we think of such assumptions as attempting to fix what the correct theory or theories of reference say about a particular case. Consider the quasi-Chomskian account of the project reconstructing a theory of reference that we discussed above. On the quasi-Chomskian account, theories of reference are attempts to reconstruct an underlying implicit theory of reference. And arguments from reference would be compelling because we take them to be based (in part) on an account of what this theory is. But when we consider the great variation in the use of theories of reference, one starts to wonder whether the independent "signal" of the implicit theory of reference makes any difference to assumptions philosophers are making.

Of course, it may be simply that the world is more complex than the quasi-Chomskian imagines. One possibility, explored by my colleagues and I elsewhere (Mallon et al. ms.,

10/1/07

Machery et al. 2004), is that referential intuitions themselves vary systematically. In such a case, these intuitions could still be an independent basis for selecting a theory of reference and its attendant assumptions, but diversity in these theories and assumptions might be expected since different persons might be using different intuitional data.

Consider how this could apply to the metaphysics of race. There, perhaps variation in analyses of the concept of race reflects differences in the structure of different individuals' concepts. Then, individual theorists like Zack or Hardimon might be offering reasons for their selections that were independent in the relevant sense (not simply determined by what they wanted the reference relation to be), but nonetheless because there is variation among individuals, there is variation in the analyses.

Moreover, the quasi-Chomskian interpretation is only one account of how the selection of a theory of reference could have an independent basis. There could be other independent bases, other ways of determining the true theory of reference. So the variation in the use of theories of reference in arguments from reference might be the result of the use of *many different* independent bases.<sup>20</sup>

On the other hand, as we mentioned above, it could be that, as a matter of fact, philosophers choose a theory of reference and its assumptions dependently – they choose them precisely because they entail a desired outcome. In that case, the variation in the selection of theories and assumptions is, in fact, dependent upon variation in what the philosophers want the outcomes to be. If this is the case, then we are owed reasons to think that many of the assumptions philosophers make track any independent truth about

---

<sup>20</sup> It's part of the difficulty in assessing such arguments that it is often far from clear just how to decide what makes a theory of reference correct (Stich 1996).

10/1/07

the reference of the terms in question.<sup>21</sup> In the absence of such reasons, we could conclude that the widespread methodology of arguments from reference is typically based on nothing more than wishful thinking about reference.

The worry that philosophers are selecting assumptions solely on the basis of what they want them to be might be merely theoretical if not for the growing empirical evidence that motivation plays a role in forming beliefs and constructing justifications. One body of empirical evidence comes from the literature on "motivated cognition" in social psychology, which is now quite extensive.<sup>22</sup> Ziva Kunda and Lisa Sinclair (1999) summarize relevant features of the work:

A general consensus has emerged that motivation can and does color judgment ... [People] may not realize that their very process of justification construction can be biased by their motives; their justifications may draw on biased subsets of the relevant concepts, beliefs, and rules that they have at their disposal. As a result, people are especially likely to draw those conclusions that they are motivated to reach. (1999, 13)

To choose just one example from this literature, consider an experiment from Thomas Gilovich's (1983) study of biased evaluation among sports fans. Gilovich asked a group of 64 basketball fans following a championship NCAA game between UCLA and Louisville to predict the outcome of a rematch between the two teams. The actual championship game was marked by a surprise missed shot and a contentious referee call in the final two minutes of the game that may well have altered the outcome – victory for Louisville. Gilovich found that participants that initially rooted for UCLA were far more likely to predict a subsequent victory for UCLA in a rematch if they were reminded of

---

<sup>21</sup> This possibility is raised in Mallon et al. (ms), and I am especially indebted to Edouard Machery for discussion about this possibility.

<sup>22</sup> Here, I follow Nichols (forthcoming) in being motivated towards a quantitative approach by this growing body of empirical evidence.

10/1/07

this missed shot and call, than if they were not so reminded. In contrast, there was no significant difference for Louisville fans. What this suggests is that UCLA fans recruited the incident to explain the "fluke" loss of their team, and preserve their judgment of UCLA's superiority. In contrast, Louisville fans considered the play to be irrelevant to the outcome. So the groups differentially recruited this fact to support their desired outcome, illustrating the phenomena Kunda and Sinclair describe.

While Kunda and Sinclair describe a broad consensus in social psychology, their description could also serve as a summary of the emerging research program on "self-serving biases" in the quite different field of behavioral economics. In a prominent study from this field, Linda Babcock, Xianghong Wang, and George Loewenstein (1996) explored self-serving biases among unionized teachers and school boards in Pennsylvania school districts. Confidential surveys were sent to the lead contract negotiators (both union and management) in every school district in the state. Selecting 75 matched pairs in which both union and management negotiators returned the surveys allowed comparison of self-serving beliefs among negotiators. In particular, Babcock et al. asked respondents "to list school districts that were comparable to their own for the purpose of contract negotiations" (10). Because such comparisons play an important role in salary negotiations, each side has a self-serving interest in selecting comparable school districts with salaries that are either higher (for the union) or lower (for the school boards), and indeed, they found that the two groups of respondents returned different (though overlapping) lists of comparable districts, with the average income among the unions' selections being \$711 (or 2.4%) higher than that of the school boards'. This difference is

significant, both statistically and practically. In actual strike situations, the difference between union and management final offers was typically around 1% (Babcock et al. 11)!

It is worth noting with regard to both of these studies that while trying to negotiate an actual consensus (e.g. who would win a rematch or what the contracted salaries should be), participants may have an incentive to distort their selection of facts for strategic gain. But in both of these cases, there were no such incentives since subjects were responding to confidential academic studies.<sup>23</sup> Thus, the studies seem to show that self-serving biases actually alter beliefs (about who would win a rematch and about appropriate comparables) that govern responses in cases in which no direct benefit is to be had.<sup>24</sup>

This sort of empirical evidence gives us good reason to believe that motivation gives rise to real differences in belief that subsequently give rise to real behaviors with significant consequences. It also provides empirical support for our present concern that the selection of assumptions (e.g. assumptions about reference) might be influenced by the aim to accommodate a desired outcome.<sup>25</sup>

### **3. Empirically Exploring the Independence of Assumptions from Reference**

While we worry that assumptions about reference might be chosen dependently, we nonetheless might wonder what choice we have other than interpreting an author at her

---

<sup>23</sup> And these beliefs apparently have real world consequences. Gilovich found that UCLA fans who explained UCLA's loss due to the fluke were willing to bet much more on victory in a subsequent rematch than those who did not (nearly as much as Louisville fans). And in an analysis of strike propensity during the 80's, Babcock et al. find that "a district in which the union's list is \$1000 greater than the board's list will have a 49 percent higher strike rate than a district in which the average salaries of the union's and board's lists are the same" (1996, 13).

<sup>24</sup> Jolls et al. 1998, p. 1502, make this point about Babcock et al.

<sup>25</sup> Of course, the problem is not motivation per se. Arguably *all* arguments result from motivation. The problem emerges when motivation interferes with tracking the truth.

10/1/07

word. If an author is unclear about the basis of her selection of a reference relation, or, if she states or implies that she takes the theory of reference to be independently motivated by, say, an independently produced intuition, don't we simply have to take that argument on its face? What other evidence could be relevant?

There may be no way to tell, for an individual author, whether or not that author has reached a desired conclusion on independent grounds or has simply selectively recruited judgments to produce what seems to be an independently justified argument. But my present purpose is not to criticize any particular author, but to question whether components of the putatively independently selected first premise of arguments from reference really are independently selected. To do this, what we need is a way of analyzing the typically employment of such arguments that need not imply, of any particular author, that he has failed to select a theory of reference on independent grounds.

In fact, there is such a way, at least in principle. For what we really want is to know, in general, whether we ought to be suspicious of the putative independence of theories of reference, and this is a general fact (a generalization) not tied to any particular individual. If we could, for a whole population of philosophers producing such arguments (a) assess what each author wants to be the case (*the desired conclusion*) and (b) assess whether their choice of assumptions of theories of reference supports the desired conclusion (*accommodation of the desired conclusion*), we could consider the strength of the correlation. For example, if putative arguments from reference *always* entail the desired conclusion, then that would be good grounds for believing the assumptions of these arguments from reference are not really, in general, independent of

10/1/07

the desired conclusion. It would simply be too great a coincidence if these putatively independent grounds of argument always converged on a desired conclusion.<sup>26</sup>

The application of this sort of statistical reasoning has been used in other fields (e.g. Moretti 2003; Jolls et al. 1998), but has not found much of a home in philosophy until Nichols's recent push for a “quantitative history” of philosophy. Nichols hopes that by considering correlations between elements of philosophical positions statistically, he can sidestep questions about the interpretation of individual authors and get at the general question of whether the philosophers embracing a philosophical position (in his case, compatibilism) are engaged in wishful thinking. While Nichols applies his strategy to historical figures, in principle such a strategy could be applied to contemporary figures as well. Thinking about our own issue in this way involves treating philosophers and their philosophical views as part of the causal order, and treating correlations between elements of a philosophical position as evidence of causation.

### **3.1. Proto-Empirical Study 1: The Dearth of Evidence**

My suspicions raised about the independence of arguments from reference, and armed with Nichols's suggestion of a quantitative approach to philosophical questions, I resolved to pilot a study into the independence of invocations of reference in the philosophical literature. I selected a set of leading philosophical journals, a starting date,

---

<sup>26</sup> It is worth noting again, that even a perfect correlation of outcomes of (putatively independent) arguments from reference and desired conclusions would not show, for any individual author, that selection of referential assumptions was is not truly independent. If arguments from reference truly have independent bases, they would sometimes converge on and sometimes diverge from desired conclusions. So the strategy here is not about interpreting any individual philosopher, it is about interpreting the general worth of invocations of theories of reference in putatively independent arguments from reference.

10/1/07

and resolved to go through each article of each issue for five years noting, for each article, (i) whether there was an (implicit or explicit) argument from reference, (ii) what the author or authors' desired conclusion (if any) was, and (iii) whether the conclusion of that argument accommodated the desired conclusion.

My pilot did not get very far, however. While it was somewhat difficult to identify clear cases of arguments from reference (thus introducing real concerns about interpretive bias), the deeper problem was that there typically was no way to identify what an author desired a conclusion to be because there was generally no explicit indication of authors' desired outcomes.<sup>27</sup> Often, authors simply do not discuss what they want the conclusion of an argument to be. Moreover, there is a very wide range of possible motivations for wanting a conclusion to be true that might lead to the kind of selective choice of assumptions that concerns us here. Among these possible motivations are:

1. A conclusion coheres with an author's unreflective intuition.
2. A conclusion is surprising. It defies folk belief.
3. A conclusion is not surprising. It fits well with folk belief.
4. A conclusion fits well with other things the author has written or believed.
5. A conclusion has beneficial political implications.
6. A conclusion will allow publication or further the author's career.

Because there is often little or no interpretive evidence that could offer a good guide to what a desired conclusion would be, there is no basis for pursuing a general study of the correlation of outcomes of putatively independent arguments from reference and desired

---

<sup>27</sup> The problem of experimenter bias is a real one, both here, and in the discussion of race below. In this way, the current discussion falls short of the much cleaner approach employed by Nichols in his own quantitative approach.

10/1/07

conclusions. So while I continued to have doubts about the independence of assumptions about reference in philosophical discussions, I had no quantitative evidence to back these doubts up.

### **3.2. Proto-Empirical Study 2: The Case of Race**

But we have already mentioned one area of philosophical debate where there are independent grounds for inferring an author's desired conclusion: race theory. In the philosophy of race, there have been two parallel discussions over the last two decades: the one we have already discussed concerning the metaphysics of race and a second one concerning the normative (including political, social, and moral) significance of "race" talk. By "race" talk, I mean the use of terms like "Asian," "black," "white," "Latino," "Native American," and their associated concepts to label and differentially treat people. For reasons we consider below, it is plausible to think the normative discussion about "race" talk gives information about desired conclusions in the metaphysical debate.

As we discussed above, with regard to the metaphysical question, three positions have been salient: skepticism, naturalism, and constructionism. Skeptics hold that race does not exist. Naturalists hold that race does exist and is a biological kind. And constructionists hold that race does exist, but is the "socially constructed" product of our social practices. What we did not discuss is the parallel normative debate about racial classification. The central questions motivating interest in race are normative, driven primarily by the moral, social and political questions surrounding whether and to what extent racial classification and identification ought to continue. While there are many intermediate positions in this debate, we can identify two competing tendencies at the

10/1/07

poles of this debate. On the one hand, eliminativists want to reduce or eliminate the uses of racial classification (of oneself and others), typically because they view racial classification as oppressive.<sup>28</sup> On the other hand, conservationists want to retain or conserve practices of racial classification, typically because they view racial classification as helpful in combating oppression, but in some cases because they view it as valuable *tout court* (e.g. Outlaw 1995, 1996).

Setting aside the metaphysical naturalists about race, there is a pragmatic alignment between positions in the metaphysical and normative debates, and, indeed, it is common to see the questions each raises treated together.<sup>29</sup> It would be best for eliminativists if "race" talk were not merely undesirable or oppressive, but false. This is because typically, if a statement is false (for example, in virtue of containing a nonreferring term), it provides a *ceteris paribus* reason not to employ it.<sup>30</sup> Similarly, it would be best for conservationists if "race" talk were true, if there really were race. Conserving "race" talk makes more sense if we can make literally true statements in talking about race. This pragmatic alignment of normatively conservationist approaches to "race" talk with metaphysically constructionist accounts of race is well known to race theorists. Indeed,

---

<sup>28</sup> In separating the metaphysical from the normative debates, I have used the term "racial skeptic" for the metaphysical view that race does not exist and "racial eliminativism" for the normative view that "race" talk ought to be eliminated. "Eliminativism" in the racial context, on this terminology, does not label a metaphysical view about race but a normative view about "race" talk. In this way it contrasts with the standard use of "eliminativism" in the philosophy of mind (which is committed to the metaphysical view that propositional attitudes do not exist) which I employed above.

<sup>29</sup> We set naturalists aside in this discussion not because they are counterexamples to the correlation discussed here, but because evidence for or against correlation is, for them, difficult to establish for the reasons discussed in the previous section.

<sup>30</sup> As J. Angelo Corlett (2003) writes, "Considerations of justice, whether distributive or retributive, are based on assumptions that certain names have referents. Where they do not, then policies based on such arguments can be dismissed as being implausible" (41). Cf. Glasgow (2006).

10/1/07

constructionist theorists have explicitly noted the need for a metaphysical account of race that can underwrite the progressive project of race theory. For example, Charles Mills has written of the need to "make a plausible social ontology neither essentialist, innate, nor transhistorical, but real enough for all that" (1998, xiv). Driving this ontological project is a pragmatic judgment about the best means to accomplish progressive political goals. As Paul Taylor put it, "we're better able to right the wrongs of classical racialism if we hold on to race-thinking" (Taylor 2004, 126), with the thought being that providing a constructionist social ontology supports "race-thinking" while undermining naturalistic accounts of race.

Race theorists thus recognize that the normative debate provides grounds for preferring one or another metaphysical position to be correct. Above, however, we saw that race theorists have employed (putatively independent) arguments from reference for their metaphysical conclusions. So, while it is in general difficult or impossible to compare the outcome of arguments from reference with a desired conclusion, it is possible to do this in the race debate.

Here is how. If the use of arguments from reference is driven by the desirability of their conclusions, then the metaphysical conclusions of these arguments should typically align with the normative position of the author. So, if the conclusions of arguments from reference show a fortuitous correlation with individual theorists' normative preferences, it will be evidence that the assumptions of these arguments were made to ensure those conclusions.

There are, however, assumptions in arguments from reference other than those concerned with the right substantive theory of reference. Implicit in the second step of

10/1/07

arguments from reference (in which claims about the reference of a term or class of terms is made) are assumptions about the way the world is. For example, in the eliminativist argument, the eliminativist assumes (or argues for the assumption) that the world is such as to make the theory of folk psychology seriously mistaken. In the case of race, however, the differences in outcomes of arguments from reference are traceable to differences in assumptions about the referent of racial terms. Elsewhere (Mallon 2006), I have argued that there is little disagreement in the philosophy of race about the empirical facts surrounding race. There is, for example, little disagreement about the genetic facts, the facts of human history and reproductive isolation, the facts of social life and social history and so forth. Rather, the debate is concerned with the question, "Given the facts, what counts as the referent of "race," if anything?" If that is right, then disagreement about the conclusions of arguments from reference centers upon disagreements about the right complete theory of reference to employ with regard to race thought and talk. So if the conclusions of arguments from reference fortuitously align with the normative positions of individual authors, this will be evidence that assumptions about reference were selected in order to ensure those conclusions.

And when we consider how the conclusions of these arguments compare to individual theorists' normative preferences, a striking alignment is apparent (see Table 1).<sup>31</sup> While interpreting individual authors can be difficult and error prone, the general pattern is hard

---

<sup>31</sup> Ideally, the sampled theorists would be randomly selected, or selected by some objective criterion (e.g. as in Nichols forthcoming). They were not (because I saw no way to do it). However, the theorists listed here represent a significant proportion of all the theorists making arguments from reference in the context of race, so I claim that they are representative of the debate over the metaphysics of race.

to miss.<sup>32</sup>

|                        | Metaphysical Result of Putative Argument from Reference | Normative View  |
|------------------------|---|-----------------|
| Mills (1998)           | Constructionist   | Conservationist |
| Taylor (2000)          | Constructionist   | Conservationist |
| Haslanger (2003, 2005) | Constructionist   | Conservationist |
| Hardimon (2003)        | Constructionist   | Conservationist |
| Appiah (1995, 1996)    | Skeptic   | Eliminativist   |
| Zack (1993, 2002)      | Skeptic   | Eliminativist   |
| Blum (2002)            | Skeptic   | Eliminativist   |

TABLE 1: Metaphysical Results and Normative Views

In race theory, eliminativists about race are racial skeptics, and conservationists about "race" talk are some sort of constructionist. This correlation is not the result of an entailment between these views. It is not contradictory to hold that while "race" talk is oppressive and should be discontinued, strictly speaking race does exist. Nor is it contradictory to hold that while race, strictly speaking, does not exist, it is nonetheless progressively useful to talk as if it does. But while there is no entailment between normative and metaphysical positions, the correlation between them is no coincidence either. Indeed, the most obvious interpretation of the alignment is exactly the one we were worried about above: choices about theories of reference and auxiliary assumptions are being driven by desired conclusions. In this case, normative (political, social and moral) considerations give rise to a desire for a particular metaphysical outcome, one that is subsequently supported by the selective choice of assumptions about reference.

---

<sup>32</sup> For more careful interpretations of Appiah, Mills, Taylor, and Zack, see Mallon (2006).

10/1/07

As with the studies of sports fans and school contract negotiators we considered above, this study is not experimental, and as such it is difficult to establish causation between our two factors with certainty. It is, for example, consistent with the correlation I have identified to insist that causality goes the other way.<sup>33</sup> Nonetheless, a number of considerations count against this suggestion. First, the suggestion is relatively implausible given the absence of a causal story connecting the outcome of the metaphysical argument to the progression of the normative argument. My suggestion is that the normative debate drives the desire for a particular metaphysical outcome, producing the selection of assumptions about reference that accommodate the desired conclusion. But it is harder to see how a metaphysical view about whether or not race exists as a social construction determines an answer to the normative argument. Second, this alternative reading seems at odds with the explicitly instrumental approach to the metaphysics taken by theorists within the literature, most obviously by constructionists.<sup>34</sup> As noted above, constructionist conservationists explicitly motivate the metaphysical project by reference to the normative one. And it is not hard to read skeptical eliminativists in the same way. Indeed, the alignment between the conclusions of arguments from reference and the normative outcomes has led some within race theory to suspect the value of arguments from reference precisely on the grounds that the arguments

---

<sup>33</sup> It also might be that the effect results from a common cause or that it is some sort of selection effect.

<sup>34</sup> Philosophers of biology also sometimes suspect that political aims are driving the selection of assumptions about the reference of "race" (e.g. Pigliucci and Kaplan 2003, 1165-66).

10/1/07

seem to be driven by the political, social and ethical considerations in the area, and are, therefore, superfluous.<sup>35</sup>

Returning to our general concern, the situation in race theory allows us at least to say this: if the authors above are, for the most part, interpreted correctly, and if we have correctly identified the cause of the correlation between metaphysical and normative positions regarding race and "race" talk, then we have found an area of philosophical discourse in which precisely the concern identified in **Section 2** is realized. Theories of reference and their accompanying assumptions are ostensibly invoked to make a difference, but here they do not. This ought, at the very least, to make us take the concern seriously.

Can we infer anything further from this fact? In particular, can we induce from our sample population of philosophers to philosophers as a whole? Can we conclude from the fact that in one area in which we can critically examine the independence of assumptions about reference, such assumptions are not, typically, independently selected, that in other areas we have reason to doubt whether assumptions about reference are independently chosen?

### **3.3. From the Metaphysics of Race to the Rest of Philosophy**

There are many problems with inferring from the case of race theory to more general philosophical practice. Our sample size is small. The interpretation of some the texts I have offered might be disputed and offers substantial room for experimenter bias. And

---

<sup>35</sup> Paul Taylor (2000) makes this point forcefully, claiming that the metaphysical debate was such that, "Appiah's very real and important ethical concerns" are hiding "in the shadow of metaphysical speculations" – speculations framed as arguments from reference – about the nonexistence of race (p. 126).

10/1/07

crucially, even if these problems could be fixed, there would remain a question about whether the way a population of race theorists employ arguments from reference is representative of the population of philosophers at large. In particular, race theory is of interest because of its political and social significance, and theorists who take part in race theory typically do so, in part, because they have strong political and social leanings. It is precisely this fact that makes it possible to glean what particular theorists hope an outcome will be. But it would also not be surprising if such strong political and social sentiments influenced the course of arguments from reference. In many other areas of philosophy, such political and social leanings are much less significant, and so perhaps they pose a much smaller risk of distorting the independence of arguments from reference.

This argument seems to me correct in that it identifies a potentially important difference between our sample and the population of all philosophers using theories of reference. And unfortunately, it raises the specter that any philosophical area in which we have systematic evidence of desired conclusions might be an area in which that fact alone gives evidence for the unrepresentativeness of work in the area. So much the worse for applying quantitative thinking to investigating the motivated choice of philosophical assumptions, perhaps.

Still, while it is true that many other areas of philosophy are not structured by strong political and social concerns, it would be wrong to conclude that other strong motivations are not playing a role. The difference between philosophy of race and these other areas may be less that philosophers of race have desired conclusions than simply that their desired conclusions fall along an axis of public concern and so are readily declared and

10/1/07

identified. In contrast, other philosophical domains may be characterized by many, disparate individual motivations. Our sample of philosophers of race is unrepresentative in virtue of their having motivations only if other philosophers do not.

At a minimum, the lesson to draw from the philosophy of race is that the worry raised above that seemingly independent arguments from reference may in fact be produced in order to accommodate desired conclusions is sometimes realized in actual philosophical discussion. And more substantially, if the correlation in the philosophy of race is indicative of a more general trend in using putatively independent arguments from reference, then the fact that we cannot detect these correlations elsewhere ought to give us little confidence that they don't exist.

#### **4. Independence and Motivation**

Suppose what I have suggested is in fact correct, and it is the case that the assumptions made about reference in typical arguments from reference are gerrymandered to produce desirable conclusions. Then it would turn out that those who produced the arguments did so on the basis of bad reasons. And it is also the case, if what I said in Section 2.3 is correct, that it is not obvious what independent reasons there might be, especially for the many auxiliary assumptions needed to apply theories of reference to particular cases. So I close by reflecting on two different paths one might pursue in attempting to respond to worries about the dependence of arguments from reference.

The first path is to put the selection of assumptions about reference on firmer ground by offering reasons for the assumptions and evidence for their independence. We ought to be suspicious of any attempt by a theorist to insist on the independence of their own

10/1/07

reasons for making particular assumptions. For example, we ought to be suspicious of attempts to insist that one really has an independent referential intuition of a certain sort. To begin with, there is no reason to think people have introspective access to the processes that produce their intuitions, and so there is no reason to think a person can judge whether or not their own intuition is independent in the right way.<sup>36</sup> Moreover, given the examples of discourse about the metaphysics of race together with the literature on motivated cognition, we might reasonably conclude that people may be mistaken about why they have arrived at the conclusion they have.

But there is another way to proceed here, consistent with insisting on the importance of intuitions, which is to explore referential intuitions experimentally. If some version of the quasi-Chomskian story is right, careful experimental investigation of referential intuitions ought to provide an independent basis for saying what the right assumptions to make in a complete theory of reference are. This is, of course, one strand of the contemporary "experimental philosophy" movement, and with luck, it might eventually provide just the sort of data that would show one or another complete account of reference best captures folk intuitions.<sup>37</sup> My colleagues and I have elsewhere suggested grounds for doubting this approach will be successful for vindicating arguments from reference (Mallon et al. ms.), but it would address the worry that our choices about reference are not independently constrained.

There is a second path, however, available in cases where the desires that I have suggested are driving discussions about reference are directed toward matters of mutual,

---

<sup>36</sup> The point about introspective access is an old one, amply documented in Nisbett and Wilson's famous (1977) paper "Telling More Than We Can Know."

<sup>37</sup> For example, in the context of race, Joshua Glasgow has been experimentally exploring folk judgments about racial classification.

10/1/07

public concern.<sup>38</sup> In our case study of race, for example, I have suggested that arguments from reference have been driven by political, social, and moral concerns about "race" talk. In such cases, we could choose to abandon arguments from reference in favor of pragmatic, evaluative debate about the worth of particular ways of describing the world (e.g. "beliefs don't exist," or "race does exist") without a detour through the theory of reference.<sup>39</sup> While this kind of debate does not offer the seemingly definitive outcomes of arguments from reference, it also does not make philosophically and practically significant questions hostage to the questions "what is the right theory of reference?" and "what are the right auxiliary assumptions" – questions whose method of resolution is far from clear, and whose prospects for actual resolution seem quite poor.

One purpose of this discussion has been to raise a question about exactly what role discussions of reference are to play in philosophical discussions outside the philosophy of language. My worry has been that assumptions about reference are hijacked by the desire for particular outcomes from arguments from reference – a worry that is consistent with what we found in our case study of the metaphysics of race. While showing that someone made an assumption for a bad reason does not show that there are not good reasons for making the same assumption, in the case of the auxiliary assumptions about reference (that can alter the philosophically significant conclusions of such arguments), it is hard to know what these good reasons could be. And without such reasons, why should we take arguments from reference seriously?

---

<sup>38</sup> It would not work as well in cases where the concerns that drive us to select assumptions are personal, e.g. concerns about our careers or individual aesthetic preferences.

<sup>39</sup> This is consistent with a strategic approach to the metaphysics of race of the sort suggested in Haslanger 2000, Shelby 2002, and Mallon 2006.

## Bibliography

- Adams, R. M. (1979). "Divine Command Metaethics Modified Again." Journal of Religious Ethics 7(1): 66-79.
- Andreasen, R. (2005). "The Meaning of 'Race': Folk Conceptions and the New Biology of Race." Journal of Philosophy CII(2): 94-106.
- Andreasen, R. 1998. A New Perspective on the Race Debate. *British Journal of the Philosophy of Science* 49:199-225
- Andreasen, R. 2000. Race: Biological Reality of Social Construct? *Philosophy of Science* 67 (Proceedings). S653-S666.
- Appiah, K. A. (1995). The Uncompleted Argument: Du Bois and the Illusion of Race. Overcoming Racism and Sexism. L. A. Bell and D. Blumenfeld. Lanham, MD, Rowman and Littlefield: 59-77.
- Appiah, K. A. 1996. Race, Culture, Identity: Misunderstood Connections. . In *Color Conscious: The Political Morality of Race*, ed. K. A. Appiah, A. Guttman, pp. 192. Princeton, NJ: Princeton University Press.
- Babcock, L., X. Wang, et al. (1996). "Choosing the Wrong Pond: Social Comparisons in Negotiations that Reflect a Self-Serving Bias." The Quarterly Journal of Economics 111(1): 1-19.
- Bishop, M. and S. P. Stich (1998). "The Flight to Reference, or How Not to Make Progress in the Philosophy of Science." Philosophy of Science 65: 33-49.
- Blum, L. (2002). I'm not a Racist But... Ithaca, NY, Cornell University Press.
- Boyd, R. (1988). How to Be a Moral Realist. Essays on Moral Realism. G. Sayre-McCord. Ithaca, NY, Cornell University Press: 181-228.
- Churchland, P. (1981). "Eliminative Materialism and the Propositional Attitudes." Journal of Philosophy LXXVII(2): 67-90.
- Corlett, J. A. (2004). Race, Racism, and Reparations. Ithaca, NY, Cornell University Press.
- Evans, G. (1973). "The Causal Theory of Names." Supplementary Proceedings of the Aristotelian Society 47: 187-208.
- Field, H. (1986). The deflationary concept of truth. Fact, Science, and Value. G. MacDonald and C. Wright. Oxford, Blackwell: 55-117.

10/1/07

Field, H. (1994). "Deflationist Views of Meaning and Content." Mind **103**: 249-285.

Gilovich, T. (1983). "Biased Evaluation and Persistence in Gambling." Journal of Personality and Social Psychology **44**(6): 1110-1126.

Glasgow, J. (2003). "On the New Biology of Race." The Journal of Philosophy **September**(456-474).

Glasgow, J. (2006). "A Third Way in the Race Debate." Journal of Political Philosophy **14**(2): 163-185.

Hardimon, M. (2003). "The Ordinary Concept of Race." Journal of Philosophy **C**(9): 437-455.

Haslanger, S. (2000). "Gender and Race: (What) Are They? (What) Do We Want Them To Be?" Noûs **34**(1): 31-55.

Haslanger, S. (2003). Social Construction: The "Debunking" Project. Socializing Metaphysics: The Nature of Social Reality. F. Schmitt. Lanham, MD, Rowman and Littlefield: 301-325.

Haslanger, S. (2005). "What Are We Talking About? The Semantics and Politics of Social Kinds." Hypatia **20**(4): 10-26.

Hieronymi, P. (2005). "The Wrong Kind of Reason." Journal of Philosophy **102**(9): 437-457.

Horgan, T. and G. Graham (1990). "In defense of southern fundamentalism." Philosophical Studies(62).

Horwich, P. (1994). Truth. Oxford, Blackwell.

Jolls, C., C. R. Sunstein, et al. (1998). "A Behavioral Approach to Law and Economics." Stanford Law Review **50**(5): 1471-1550.

Kitcher, Philip. 1999. Race, Ethnicity, Biology, Culture. In Harris (1999). 87-120.

Kripke, S. 1972. *Naming and Necessity*. Harvard University Press. Cambridge, MA.

Kunda, Z. and L. Sinclair (1999). "Motivated Reasoning With Stereotypes: Activation, Application, Inhibition." Psychological Inquiry **10**(1): 12-22.

Lewis, D. 1970. How to define theoretical terms. *Journal of Philosophy*. 67:426-446.

Lewis, D. (1972). Psychophysical and theoretical identifications. *Australasian Journal of Philosophy*, 50, 249-58.

10/1/07

- Lycan, W. G., 1988. *Judgement and Justification*. Cambridge: Cambridge University Press.
- Machery E., Mallon R., Nichols S, and Stich, S. 2004. "Semantics, Cross-Cultural Style." *Cognition*. 92: B1-B12.
- Mallon, R. (2006). "'Race': Normative, Not Metaphysical or Semantic." *Ethics* **116**(3): 525-551.
- Mallon, R. (2004). "Passing, Traveling, and Reality: Social Construction and the Metaphysics of Race." *Noûs* **38**(4): 644-673.
- Mallon R., Machery E., Nichols S, and Stich, S. ms. "Against Arguments from Reference."
- Mallon, R. and S. P. Stich (2000). "The Odd Couple: The Compatibility of Social Construction and Evolutionary Psychology." *Philosophy of Science* **67**: 133-154.
- Moretti, Franco. 2003. "Graphs, Maps, Trees: Abstract Models for Literary History – 1," *New Left Review* 24: 67-93.
- Nichols, S. (forthcoming). "The Rise of Compatibilism: A Case Study in the Quantitative History of Philosophy." *Midwest Studies in Philosophy*.
- Nisbett, R. E. and T. D. Wilson (1977). "Telling more than we can know: Verbal reports on mental processes." *Psychological Review* **8**: 231-259.
- Outlaw, L. 1995. On W.E.B. Du Bois's "The Conservation of Races". In L. A. Bell, D. Blumenfeld (1995). 79-102.
- Outlaw, L. 1996. *On Race and Philosophy* New York: Routledge.
- Pigliucci, M. and J. Kaplan. (2003). "On the Concept of Biological Race and Its Applicability to Humans." *Philosophy of Science* **70**: 1161-1172.
- Putnam, H. (1975). The meaning of 'meaning'. *Mind, Language and Reality: Philosophical Papers*. New York, Cambridge University Press. **2**: 215-271.
- Shelby, T. (2002). "Foundations of Black Solidarity: Collective Identity or Common Oppression." *Ethics* **112**.
- Smith, M. (1994). *The Moral Problem*. Oxford, Blackwell.
- Stich, S. (1996). Deconstructing the Mind. *Deconstructing the Mind*. New York, Oxford University Press: 3-90.

10/1/07

Stich, S. P. (1983). From Folk Psychology to Cognitive Science: The Case Against Belief, MIT Press. A Bradford Book.

Taylor, P. (2000). "Appiah's Uncompleted Argument: Du Bois and the Reality of Race." Social Theory and Practice **26**(1): 103-28.

Taylor, P. C. (2004). Race : a philosophical introduction. Malden, MA, Polity Press.

Tye, M. (1998). "Externalism and Memory." The Aristotelian Society **Supp Vol.**

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

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