

**'Good' On Twin Earth** <sup>1</sup>

by  
Geoffrey Sayre-McCord  
UNC/Chapel Hill

Introduction: Moorean Shenanigans

Wanna have some fun? Start a moral conversation with a sincere, thoughtful, reflective acquaintance and then, at an appropriate moment, ask her what she means by the moral terms she throws around so freely. Of course, if she has her wits about her, she might just respond -- indignantly -- by saying, "You know perfectly well what I mean" ... at which point the fun pretty much stops, because she's right and if you're decent you'll admit it. However, if your timing is good, and if she has a philosophical bent, she will take you seriously. She might then mention a bunch of things (happiness, love, honesty,) that she thinks are good, say, or point to the fact that people normally feel guilty when they do something wrong, or perhaps highlight the connection between justice and equality.

With the lessons of G. E. Moore behind you, however, it won't take long to convince her that she doesn't have much of an answer. It is really not hard to bring home Moore's point that there's a difference between the things that have some property and the property itself, nor is it hard to highlight the difference between some property considered in itself and the relations (causal and otherwise) that it might bear to other things. Sure happiness, love, and honesty are good, but (we might press) the question is: what are we saying when we say they are all good -- what property is it that we take them to share? And sure, doing something wrong often causes feelings of guilt, and justice may require that we treat people equally, but (we might press) in virtue of what do the things that (often) cause guilt count as wrong and what is the nature of justice such that it demands equal treatment?

The fun, fairly had, comes at our own expense, of course, since whatever answers we philosophers might venture don't seem much better. Moore, in fact, insisted that the only good answer anyone has is the indignant one our acquaintance might give: we all damn well know what "good" and "right" and "just" mean! ... "good is good," he pointed out, "and that is the end of the matter...if I am asked 'How is good to be defined?' my answer is that it

cannot be defined, and that is all I have to say about it."<sup>2</sup> Famously, Moore himself drew radical conclusions from this concerning the metaphysics of morals. In particular, he held that our moral terms, given that they are meaningful, must refer to non-natural properties that depend in some way on, but are not one with, the natural properties discoverable by sense and science. I have worries about the sort of metaphysics Moore advanced, but I think Moore was on the right track, more or less, in holding that our moral terms are indefinable.

Now you might think that what I've just suggested is both cruel -- when it comes to having fun -- and overly pessimistic -- when it comes to the definitions philosophers have to offer.

I admit, having fun in the way I suggest could be cruel, depending on how one does it and with whom. But it needn't be: after all, the questions raised *are* genuinely interesting and our inability to answer them easily is more than a little surprising. If the fun is, as I put it, fairly had, then it needn't occasion any serious embarrassment or discomfort, just some tenacious puzzlement.

As for whether I am being overly pessimistic about our ability, as philosophers, to provide definitions for our own moral terms...I doubt it. In any case, as will become clear, my pessimism is highly localized -- it concerns only our ability to provide a certain kind of definition of our moral terms -- and it will come wrapped in a view that is in other regards optimistic about our ability to say interesting and informative things about how moral terms come to have the meaning they do and about the properties we use them to refer to.

What put me in mind of these Moorean shenanigans is a recent argument that has been offered concerning the semantics of moral language. The argument comes from Graceland (Memphis) thanks to Terence Horgan and Mark Timmons, and it defends noncognitivism in ethics *via* a moral twin earth argument.<sup>3</sup> While I think the argument geographically suspect (Graceland is, after all, a land of fantasy), and in the end mistaken, it represents a serious challenge to the sort of view I would want to defend -- a view that embraces cognitivism but steers clear of analytic definitions.

---

<sup>2</sup>. *Principia Ethica*, (Cambridge University Press, 1903), p. 6.

<sup>3</sup>. The argument is developed by Terence Horgan and Mark Timmons in three places: "Troubles on Moral Twin Earth: Moral Queerness Revived," *Synthese* 92 (1992) pp. 221-260; "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived," by Terence Horgan & Mark Timmons in *Philosophical Papers* XXI (1992), pp. 153-175; and in "New Wave Moral Realism Meets Moral Twin Earth," by Terence Horgan & Mark Timmons in *Journal of Philosophical Research* XVI (1990-91), pp. 447-465. Roughly the same argument can be found in Simon Blackburn's "Just Causes" in *Essays in Quasi-Realism* (Oxford University Press, 1993) and Michael Smith's *The Moral Problem* (Blackwell, 1994).

---

<sup>1</sup> Copyright 1996 by Geoffrey Sayre-McCord

My preferred view borrows heavily from contemporary philosophy of language by introducing, on analogy with natural kinds, the notion of moral kinds.<sup>4</sup> Among other things, since (on this account) the extensions of moral terms are not determined by the particular beliefs people happen to hold, the moral kinds approach helps to make sense of how people with wildly divergent moral conceptions can still be disagreeing with -- rather than talking past -- one another. Moreover, because it treats proposed 'definitions' of moral terms as discoveries not settled by current use, prevailing linguistic conventions, or stipulative declaration, the moral kinds approach also helps to make sense of why no robust analytical definitions are available.

Unlike others who have found hope for moral semantics in contemporary philosophy of language, I will, however, strenuously resist treating moral terms as natural kind terms.<sup>5</sup> Moral terms, I'll argue, do operate *much like* natural kind terms, especially in their role as putatively referring terms and in their resistance to analytical definitions. Yet they differ in one crucial respect, at least: In contrast with natural kind terms, we don't, and don't believe we should, defer to scientists or scientific theory in determining the true nature of what we are referring to in using moral terms. Instead, we adjust our views of what is good, or right, just, or obligatory, as we change our moral (and more broadly, normative) theory, not our scientific theory. The developments crucial to moral taxonomy are found squarely in normative theory and not in social or psychological theory, let alone biological or physical theory -- except as these fields influence, and are countenanced as relevant by, our normative theory.

Discovering the referents of our moral terms is a process, I'll suggest, of discovering what normatively significant kinds -- and not causal-explanatorily significant kinds -- regulate our moral beliefs. The regulation

---

<sup>4</sup> The idea that the model of natural kinds might be appropriate to moral language has been around for a while. Putnam, for instance, suggests it in "Language and Reality," in Mind, Language and Reality, (Cambridge: Cambridge University Press, 1975), p. 280; David Brink makes use of it in replying to arguments against moral realism, in Moral Realism and the Foundations of Ethics (Cambridge: Cambridge University Press, 1989); and Richard Boyd pursues it in some detail in "How to Be a Moral Realist" in Essays on Moral Realism, G. Sayre-McCord (ed.), (Ithaca: Cornell University Press, 1988), pp. 181-228. See also William Lycan's "Moral Facts and Moral Knowledge," in The Southern Journal of Philosophy, XXIV (1986), Spindel Conference Supplement, pp. 79-94.

<sup>5</sup> I don't think I am alone in this, yet even among those I think might accompany me, little attention is paid to what moral terms might refer to if not natural kinds. In discussing the possibility of ethical naturalism, David Brink notes, in a footnote, that even those who think some moral term designates a property that is designatable as well by some non-moral predicate, needn't hold that the terms in question constitute kind terms within natural science. He doesn't, however, pursue the thought. See Moral Realism and the Foundations of Ethics, fn., p. 157. The one exception might be Susan Hurley, who defends a view that, I think, is much like mine. See Natural Reasons, (Oxford University Press, 1989).

involved must be, I'll nonetheless assume, a causal regulation, so the kinds in question will end up being causally relevant at least to the explanation of our use of the terms. Yet when a moral kind is appropriately implicated in the explanation of our use of some term it will not be because we causally interact with the kind itself. Rather, it will be because (i) our use is causally responsive to what are, in fact, instances of the kind and (ii) the use to which the term is put is one of referring to whatever normatively significant kind it is that they are instances of.<sup>6</sup>

Where the contrast between moral kind terms and natural kind terms shows up is in answering two questions:

In virtue of what does the motley crew of things influencing the use of a term count as an instances of *the same (relevant) kind of thing*?

and

When are we to think of someone's words as different from ours in meaning because sensitive (that is, causally sensitive) to some other kind of thing?

A natural kinds model would see us, in answering these questions, as deferring to natural (or social) science for the relevant taxonomy of kinds; a moral kinds model won't. On the moral kinds approach, two people will be seen as using moral terms with the same meaning if, but only if, the use of the terms in question are appropriately causally regulated by the same moral, not natural, kind. Of course just what might count as appropriate causal regulation is itself a complicated and controversial matter. I assume that a crucial part of that story will highlight features of the use of the term in question that make its use responsive to new information about the kind in question. In any case, on this view, different people may be using moral terms with the same meaning -- to refer to the same relevant kind -- even though their respective use of the terms is appropriately causally regulated by events, or actions, or institutions, that fall into radically different natural kinds. What is important, if moral terms refer to moral kinds, is that this possible natural diversity reflect a moral homogeneity.

At least in principle, we might succeed in defending a particular moral theory, with its attendant taxonomy, without having any sort of naturalistic definition of the kinds it countenances. It might turn out that the moral kinds the

---

<sup>6</sup> A parallel point holds concerning natural kinds. When a natural kind is appropriately implicated in the explanation of our use of some term it will also not be because we interact with the kind itself. Rather it will be because (i) our use is causally responsive to what are, in fact, instances of the kind and (ii) the use to which the term is put is one of referring to whatever explanatorily significant kind it is that they are instances of.

theory identifies have contours that do not map onto (even quite complex) natural kinds. If naturalism is true, of course, then whatever the contours of the moral kinds are, the boundaries among moral kinds must reflect some natural difference between those things that belong to a particular moral kind and those that do not. However, the natural differences in question may not be differences in natural *kind* but only, say, differences in degree. Similarly, given the supervenience of the moral on the nonmoral, whatever the contours are of the moral kinds, their boundaries must reflect some non-moral difference between those things that belong to a particular moral kind and those that do not. Since the nonmoral differences may, for all supervenience requires, be nonnatural differences in, for instance, the effects different actions have on the soul of the agent, the constraints supervenience imposes may not line up with those imposed by naturalism. Nonetheless, doing justice to supervenience does require recognizing that there must be, at least in principle, some nonmoral difference between those things that do, and those that do not, fall within a particular moral kind. Yet the nonmoral difference may, for all that, be a significant difference only from the point of view of normative theory.

A naturalist will presumably want assurances that what falls into the appropriate kinds, and why they so fall, can be understood not just in nonmoral terms but naturalistically. However, these naturalistic desiderata are compatible with a wide variety of views about moral kinds and their status, and one might adopt the moral kinds account of the meaning of moral terms without taking any stand on naturalism.

Indeed I am hoping there is a way to describe the semantics that leaves it free from pretty much any particular metaphysical commitment. So while I am after an account that will underwrite a cognitivist interpretation of moral language, it is no part of my project here to press a metaphysical point.

In any case, I believe the disanalogy between natural and moral kind terms goes a long way towards accounting for why the attempt to find a causal-explanatory role for moral properties strikes so many not simply as hopeless but as hopelessly off-base. Whether we can find such a role for moral properties is not, I believe, *completely* irrelevant, but the relevance of such attempts depends on our independently being able to make out a theory's normative credentials.<sup>7</sup> At the same time, once the disanalogy is appreciated, I think the apparently

---

<sup>7</sup>. In various places I've attempted to ward off peremptory objections to moral realism that are grounded in holding that moral properties are either in principle -- or at least obviously -- explanatorily irrelevant. But in denigrating peremptory objections I've tried still to acknowledge that the concerns they express do have a place. See my "Moral Theory and Explanatory Impotence" in *Midwest Studies* XII (1985), pp. 433-457, and "Moral Explanations" in *Philosophical Perspectives*, Vol. VII edited by James Tomberlin (1992), pp. 55-72.

formidable objection from Graceland, which is designed to show that the new semantics doesn't fit moral terms, will be seen to miss its mark.

My main aim is to make good sense of how the moral terms we all toss around so freely might contribute substantially to the content of what we say. Or rather, that is my over-arching aim. My proximal aim -- my aim in this paper -- is less ambitious. Here I hope to defend my preferred approach to moral semantics from the challenge from Graceland. Consonant with this more modest aim I will not be doing as much as I would like to offer a positive argument for my preferred view. I won't, for instance, work through the variety of cases one would need to appeal to in order to establish that moral terms are in fact kind terms. And I will simply ignore the morass of issues that come in tow when one views attempts at justification as theorizing and normative theory as getting at the true nature of evaluative properties.<sup>8</sup> What there is by way of a positive argument here is found in the observation that the moral kinds semantics -- if it does apply to our moral terms -- would allow that moral terms contribute to the content of what say in using them, without supposing that they are amenable to analytic definitions.

#### Keeping the Right Questions Open

G. E. Moore set the problem that shapes this paper, which is to give a plausible account of the meaning of moral (and, more generally, evaluative) terms. He set it, famously, by deploying the Open Question Argument. The argument is familiar, so I will take the time here just to gesture in its direction.

As Moore would have it, if a property-referring term is meaningful, it is meaningful thanks to its expressing a concept which in turn determines the property to which the term refers.<sup>9</sup> If the term is (analytically) definable, it will be definable by terms that themselves express other concepts that taken together constitute the concept expressed by the term being defined. Thus if two property-referring terms have the same meaning and so refer to the same property, that will be because they express the same concept. If the concept they express is simple, then while the terms may in a sense be interdefinable, neither term is in Moore's sense analytically definable. If the concept they express is complex, however, then the terms are in principle definable (by terms that taken together express all the components of the complex concept), and the property the terms refer to will likewise (Moore thinks) count as complex (and so analyzable).

---

<sup>8</sup>. I do try to defend these views in "Moral Kind Terms," manuscript.

<sup>9</sup>. Actually, Moore talks of the "object or notion denoted by a word" and he seems to switch back and forth casually between talk of the notion (which he sometimes also calls an idea) expressed and the object or quality referred to (pp. 5-9).

When it comes to our evaluative terms, then, Moore sees three possibilities: (i) an evaluative term might express a simple concept, in which case the term is not analytically definable, (ii) it might express a complex concept, in which case the term is definable by the terms that express the concepts that are components of the concept, or (iii) it might express no concept at all, and so be meaningless.<sup>10</sup>

Having set out these possibilities, and focussing on the term 'good', Moore argues against the second possibility first. He holds that 'good' does not express a complex (and so analyzable) concept on the following grounds: for any proposed definition of 'good', where the definition mobilizes some complex concept, expressed by 'x', it can always be intelligibly asked "Is x good?". For example, if someone proposes that being good is a matter of being an object of a desired desire, we can intelligibly ask "Is what we desire to desire good?". That this is an open question, not settleable immediately in the minds of those competent with the relevant terms, shows (Moore holds) that 'good' and 'an object of a desired desire' express different concepts and so refer to different properties. Substitute whatever definition you please and the question will, Moore says, remain open. This fact shows (he thinks) that there is no complex concept that can legitimately be identified with the concept expressed by 'good'.<sup>11</sup>

Turning to the third possibility -- that 'good' expresses no concept and so is meaningless -- Moore points to the *intelligibility* of the various open

questions. That the questions make sense shows, he holds, that the terms in them (including the term 'good') must be meaningful.<sup>12</sup>

That leaves only the first possibility: (1) 'good' must express a simple unanalyzable concept. Moreover, as Moore sees it, the Open Question Argument shows not only that 'good' is indefinable and yet meaningful, it shows as well that the simple property it refers to is non-natural -- since for any concept, *c*, of a natural property (complex or not) it will always be an open question whether *c* is good.

Many have been convinced by the Open Question Argument that evaluative terms are not definable -- at least not in naturalistic terms -- and convinced too that they do not refer to natural properties. Yet troubled by a metaphysics of nonnatural properties and an epistemology that relied on our having an (apparently inexplicable) intuitive access to those properties, many have gone back to Moore's original trilemma and argued that despite peoples' sense that they understood the meaning of moral terms, those terms express no concept and refer to no property. As a result, the terms contribute no content to the sentences in which they occur and are (in Moore's sense) meaningless. Moral terms may have some sort of meaning, but whatever meaning they have is a matter of their playing an expressive or directive role, not a matter of their referring to properties (nonnatural or otherwise).<sup>13</sup> The main difficulty facing this sort of view is to explain why moral discourse mimics so well the behavior of meaningful, factual, discourse. But this is a difficulty many have thought can be met.<sup>14</sup>

Recently, of course, the significance of the Open Question Argument has come in for re-evaluation, thanks largely to the realization that non-synonymous terms -- terms that do not express the same concept -- may nonetheless refer to the same property, just as different names may refer to the same individual.<sup>15</sup> Thus, for instance, *being water* is one and the same with

---

<sup>10</sup>. p. 15

<sup>11</sup>. Moore did believe that some evaluative terms were analytically definable. For instance, he thought that 'right' could be analyzed roughly as 'produces the greatest amount of good'. Ross, of course, took issue with precisely this claim. Mimicking the form of Moore's Open Question argument, Ross pointed out that it makes perfect sense to ask "I know this action will produce the best consequences, but is it right?" That this question is an open question shows, Ross thought, that rightness should not be identified with producing the best consequences and that that question makes sense shows that 'right' is not meaningless. Freed of the assumption that rightness must be identified with producing goodness, Ross turned his attention to trying to discover what things had the property of rightness. He discovered that only particular actions had that quality, but that there were general rules that summarized accurately what sorts of actions tended to have the property of rightness. So, for example, he argued that acts that are instances of keeping ones promises tended to be right -- in other words, that keeping ones promises was *prima facie* right. To say of some sort of action that it was *prima facie* is just to say that most instances of that sort are right. When it came to explaining how we can know what things are right Ross followed Moore in saying that simple unanalyzable properties could be known only through inspection and intuition.

---

<sup>12</sup>. Even as Moore came to appreciate the plausibility of noncognitivism, he fell back on this consideration in resisting it. See "A Reply to My Critics," in *The Philosophy of G.E. Moore*, edited by Paul Arthur Schilpp, (Northwestern University, 1942), pp. 535-554.

<sup>13</sup>. Moore himself later acknowledged that this noncognitivist line -- which he hadn't really considered -- has significant appeal. See "A Reply to My Critics."

<sup>14</sup>. See, for instance, Simon Blackburn's "Attitudes and Contents," in *Essays in Quasi-Realism*, pp. 182-197, and Allan Gibbard's *Wise Choices, Apt Feelings* (Harvard University Press, 1990).

<sup>15</sup>. Grounds for this re-evaluation came to the fore in work by U. T. Place and J.J.C. Smart who emphasized that science provides a bountiful supply of *a posteriori* (and putatively contingent) property identities. See Place's "Is Consciousness a Brain Process?" *British Journal of Psychology*

*being composed of H<sub>2</sub>O molecules*, even though 'water' and 'composed of H<sub>2</sub>O molecules' are clearly not synonymous (as is shown, well enough, by the openness of the question: "is water H<sub>2</sub>O?"); and *being gold* is one and the same with *having an atomic weight of 79*, even though 'gold' and 'atomic weight of 79' are clearly not synonymous; and having a certain temperature is identical to having a certain mean kinetic energy, even though 'temperature' and 'mean kinetic energy' are clearly not synonymous; and so on.

The suggestion, still more recently, has been that some terms in our language (e.g. names, definite descriptions used referentially rather than attributively, and certain kind terms) refer to the individuals or properties they do not in virtue of a concept the terms express but instead thanks to their bearing an appropriate (it is usually thought *causal*) connection to the individuals or properties.<sup>16</sup> Moreover, it is commonly said, the terms rigidly designate these individuals or properties -- that is, they designate one and the same individual or property in every possible world in which the individuals or properties can be found. Thus, these terms refer to what they do even when their referents happen not to fit some description or satisfy some concept we associate with the term. That we associate some description with the term may well reflect our belief that what we are referring to satisfies the description, and we may fix the reference of our term by appeal to some such description, but our successfully referring to something does not, on this view, depend on that thing satisfying the description. We can, in effect, be seriously wrong about what we are referring to and yet still successfully refer.

The main arguments offered for thinking of some terms (say names or natural kind terms) that their reference is not determined by a description or concept with which they are associated by competent users of the term can be seen as tricked-up versions of Moore's Open Question Argument.

To see, for instance, that the reference of 'Aristotle' is not determined by a description commonly associated with the name (e.g. philosopher, student of Plato, author of the Nicomachean Ethics, teacher of Alexander), people point out that, in principle at least, we could discover that Aristotle wasn't at all as we've taken him to be; that, even if our beliefs happen to be true of him they are not necessarily true of him (Aristotle, one and the same person, could have died young, or gone in for politics instead of philosophy, or he might have

diabolically claimed for himself someone else's credentials and accomplishments); and that, in any case, not just anyone who happened to satisfy that description would, in virtue of that, be Aristotle. In other words, the question "Was Aristotle a philosopher who studied with Plato, wrote the Nicomachean Ethics...?" is in the relevant sense an open question -- it is not an analytic truth that Aristotle fits this description, nor is it true that anyone who fits this description is Aristotle.

Similarly, to see, for instance, that the reference of 'gold' is not determined by a description commonly associated with it (e.g. solid, yellow, heavy, fusible, malleable), people point out that, in principle at least, we could discover that gold isn't at all as we've taken it to be; that, even if our beliefs concerning gold happen to be true of it, they are not necessarily true of it (gold, one and the same substance, could have been different in any or all of these respects); and that, in any case, not just anything that happened to satisfy that description would, in virtue of that, be gold. In other words, the question "Is gold solid, yellow, heavy, fusible, and malleable?" is in the relevant sense an open question -- it is not an analytic truth that gold fits this description, nor is it true that anything that fits this description is gold.

Putnam's familiar and picturesque appeal to Twin Earth makes these points nicely. To argue that the reference of our term 'water' isn't determined by the description we associate with it, Putnam suggests we imagine Twin Earth where things are almost exactly as they are here, except the clear, colorless, odorless liquid that falls from the skies and fills the oceans has some complex molecular structure that makes it fundamentally different from the stuff, as it happens H<sub>2</sub>O, by which we are surrounded and of which we are largely composed. He points out that, if the stuff around us is H<sub>2</sub>O, then in discovering that the liquid on Twin Earth is a different kind of liquid we would have discovered that it is not water, however much it resembles water. And if H<sub>2</sub>O is found on Twin Earth, but it there manifests significantly different properties, we would in talking of water be talking of it despite its failing to match any of the descriptions we commonly associate with water. Moreover, if people on Twin Earth use a term 'water' to refer to the *kind* of liquid that happens to be around them, then our term and their term refer to different kinds of liquid even if we and they associate the same description with our respective terms.

Again (to force things into the open question format) the point is that it is an open question whether some liquid that is clear, colorless, odorless, falls from the skies and fills the oceans is water, and it is an open question whether water is clear, colorless, odorless, falls from the skies and fills the oceans. What closes the questions, so to speak, is not anything available to us merely in virtue of being competent with the relevant terms. In the case of natural kind terms, what closes the question is, in effect, the taxonomy of the best natural science (insofar as we take natural science to settle what natural kinds there are); for it is that taxonomy that will tell us what (relevant) *kind* of liquid it is that we are, as

---

47 (1956), pp. 44-50; and Smart's "Sensations and Brain Processes," *Philosophical Review* 48 (1959), pp. 141-56.

<sup>16</sup> Here work by Saul Kripke and Hilary Putnam is of course of central importance. See Kripke's "Naming and Necessity," in Semantics of Natural Language, ed. by D. Davidson and G. Harman (Dordrecht: Reidel, 1972) and Putnam's "Meaning and Reference," in *The Journal of Philosophy* (1973), pp. 699-711.

it happens, surrounded by; and it is that taxonomy that will determine as well whether two samples of stuff (say one from here, and one from Twin Earth) are of *the same relevant kind*.

Needless to say, two samples of liquid will be of the same kind for a number of different kinds -- they may both be of the kind "collected in a bucket" or of the kind "drinkable" or of some other kind. And they will likewise be of different kinds for a number of other kinds -- one may be of the kind "collected in the morning" while the other isn't or one may be of the kind "coveted" while the other isn't. When it comes to 'water' our willingness to hold that it does refer to H<sub>2</sub>O depends on our thinking both that natural science reveals what *relevant kind* of stuff (if any) it is that we've been referring to, and on our thinking that what science has revealed is that it is H<sub>2</sub>O. More generally, when it comes to natural kind terms, the supposition is that the taxonomy of natural science reveals the kinds that our terms refer to, if they refer at all.

In effect, then, there are two parts to this general line of thought. One part consists in arguing, concerning certain terms, that their reference is determined by something other than the concepts or descriptions we happen to associate with them. This argument is made, largely, by appeal to the tricked-up versions of the Open Question Argument which reveal our willingness to grant (i) that what we are referring to might not actually satisfy the associated concept or description, (ii) that even if it does, things could have been such that it didn't, and (iii) that something satisfying the concept or description is not sufficient to establish that we are referring to it. Satisfying the concept or description is neither necessary nor sufficient for something to count as a referent of the term.

The other part consists in a positive (if usually quite vague) suggestion as to what it is that does determine the reference of these terms, given that it is not the concept or description with which they are associated.

In the case of names, the positive suggestion is that the reference of a name is established, in the first instance, by the term being associated (*via*, say, a suitable referential intention) with the individual named, and the subsequent success of others who use the name to refer to that individual is due to their use of the name bearing some appropriate (usually causal) connection to the individual.<sup>17</sup> Thus two different names will refer to one and the same

---

<sup>17</sup>. Just what sort of connection is required is a matter of dispute, not surprisingly. Some suggest that it is maintained only if later users of the name have, among their intentions, the intention to use the name to refer to whoever was being referred to by those from whom they learned the name. Others suggest that such an intention isn't required, but that the later users of the name must nonetheless be such that their use of the name, or the beliefs they have concerning what or who is named, are appropriately causally regulated by what or who the name refers to. I am myself inclined towards the latter view, but nothing in this paper turns, I think, on the choice between these different views.

individual if the use of each term can be traced back appropriately to that individual. And the beliefs people associate either with a single name or with two names referring to the same individual, can and often will differ dramatically without severing the requisite connections.

In the case of kind terms, the positive suggestion is that the reference of a kind term is established, in the first instance, by its being associated (again, *via*, say, a suitable referential intention) with what is taken to be an instance of the kind, and the subsequent success of others who use the term to refer to that kind with the term is due to their use of the term bearing some appropriate (again, usually, causal) connection to instances of that kind. Yet, when it comes to kind terms, the story is immediately complicated by the fact that the terms refer not to this or that particular instance of the kind but to the kind of which they are instances. A distinctive feature of kind terms is that they are introduced and used against the background assumption that the particular samples bear a certain (theoretically determined) relation to one another: that of being of the same relevant kind.<sup>18</sup> Even so, two different terms will refer to one and the same kind if the use of each can be traced back appropriately to (instances of) that kind. And the beliefs people associate either with a single kind term or with two such terms referring to the same kind can and often will differ dramatically without severing the requisite connections.

The evidence marshalled in favor of these positive suggestions plays out against our willingness to grant that our terms might in fact refer to individuals or kinds that differ (or could differ) quite radically from how we take them to be, and so to our seeing that their reference is determined by something other than the concepts or descriptions we associate with them. The evidence consists largely in noticing the sorts of considerations we rely on in determining what people are referring to in using various terms. In general, we see people as talking about some thing (an individual or kind) -- even in cases where they have false beliefs about it -- as long as that thing is seen by us as implicated in the right way in explanations of why they have the beliefs they do about what they take the term to apply to.

We become convinced, for example, that two people are referring to one and the same kind of liquid by their respective uses of 'water' and 'l'eau' when we come to see that the various beliefs they hold (about the stuff they call 'water' and 'l'eau' respectively) are appropriately explained as being responsive to samples of that kind. Similarly, we might come to see, of our own use of two terms, that they refer to the same individual or kind, despite our initial conviction that they don't (or even couldn't, in light of the beliefs we rely on in

---

<sup>18</sup>. Putnam highlights this in "The Meaning of 'Meaning,'" in *Mind, Language, and Reality* (Cambridge University Press, 1975), p. 225.

applying the terms), on discovering that one and the same thing is implicated in the right way in explanations of why we have the beliefs we do about what we have applied the terms to. Lois Lane was certain that Clark Kent wasn't (and perhaps couldn't be) Superman, and for good reason given the evidence initially available to her. But she learned otherwise on learning that one and the same individual appeared to her in different guises. 'Clark Kent' and 'Superman', she learned, refer to the same individual.<sup>19</sup>

#### A Natural(ist) Strategy

Naturalists in ethics have found hope in these developments. If property identity doesn't require synonymy, then whatever the force of the Open Question Argument may be, it does not establish that our moral terms fail to refer to natural properties. Just as we have discovered (so it seems) that being water is being composed of H<sub>2</sub>O molecules, despite the fact that "Is H<sub>2</sub>O water?" is in the relevant sense an open question, so too we might discover that being good is (say) being an object of a desired desire, despite the fact that "Is an object of a desired desire good?" is in the relevant sense an open question.

A tempting strategy for naturalists in ethics has been to acknowledge (in the face of the Open Question Argument) that 'good' has no analytic definition, and then argue, first, that 'good', like 'water', rigidly designates whatever natural property causally regulates (in the appropriate way) our use of the term and, second, that some natural property or other (for instance, being an object of a desired desire) is in fact what causally regulates that use. This would put us in the position, the thought is, to defend as an *a posteriori* yet necessary truth the claim that being good is one and the same with having that property (for instance, being an object of a desired desire).

---

<sup>19</sup> It is worth noting that the resulting positive views concerning how reference is determined are not themselves offered as analytical definitions of the terms. It is no part of these views that a person is competent with the relevant terms only if she believes (or would believe if she had all the relevant concepts) that, for instance, 'water' refers to whatever science reveals to be the real nature of the stuff around us which we take to be a clear, colorless, odorless, liquid. 'Water' may have its meaning thanks to its playing a certain role -- the role of referring to a certain kind of stuff, where the relevant kind is settled by science -- without 'water' being associated by competent speakers with the concept "whatever science tells us that stuff is". Our willingness to credit someone with using a term, say 'water', to refer to the kind we refer to, will depend on our seeing their use as bearing the right relation to stuff of the same kind -- but the relation might be there even if that person resists the authority of science in determining the nature of water and even if she believes that 'water' by definition refers to 'clear, colorless, odorless liquid.' No doubt, the more resistant she is to shifting her view of what counts as water in light of the evidence we might marshal, the stronger will be our grounds for suspecting that her term refers not to that kind but instead to 'whatever is a clear, colorless, odorless liquid.' But these grounds for suspicion would have to be weighed against others we might have for thinking that the term she uses refers to water.

Of course, the general proposal needn't come with any specific commitment as to what property it is that regulates our use of the term 'good'. Just as people successfully referred to water, and believed a whole slew of true things about it, long before it was discovered to be H<sub>2</sub>O, so too we may now successfully refer to goodness, and believe a whole slew of true things about it, before we discover (if we ever do) what sort of natural property it is. Still, if one is committed both to naturalism and to there being moral properties, the semantics of natural kind terms seems to provide a safe haven in the face of the Open Question Argument. For it allows us to make sense of how moral properties might be natural properties even though moral terms are not synonymous with non-moral terms.

This sort of account has a number of other virtues as well. It does a nice job of making sense of the fact that people with dramatically different beliefs concerning value (and rightness, and virtue...) nonetheless often seem to be talking about one and the same property. It fits well too with our thinking that as our own moral views evolve and shift they might continue to be views about the same thing. And it explains easily the supervenience of the moral on the non-moral, since it identifies the properties referred to by our moral terms with properties referred to by certain non-moral terms.

Obviously, such an account of the meaning of our moral terms leaves open the possibility that when all is said and done 'good' or 'right', like 'phlogiston', will turn out to refer to no relevant kind whatsoever. Whether or not it refers depends, on this view, on whether our use of the term 'good' is causally regulated in the appropriate way by instances of some reasonably homogeneous kind that is countenanced by science (perhaps broadly construed to include social science, psychology, and biology in addition to physics). As will become clear, I think seeing the referential success of our moral terms as being hostage, in the way this proposal does, to science is a mistake, even for one with naturalistic commitments. And I think the advantages of the account can all be had without thinking of moral terms as *natural* kind terms. But let me put this point to one side, for now, and turn to the argument from Graceland. This argument is designed to show that even if some terms (say 'water', 'gold', 'heat', etc.) are well handled by the new semantics, moral terms are not.

The argument begins by recommending that we imagine a Moral Twin Earth where things are pretty much as they are here right down to the fact that people in the Twin-USA use terms orthographically identical to our 'good', 'right', 'virtuous', etc. much as we do, at least when it comes to recommending courses of action, directing praise and blame, and settling on various other attitudes. We are to imagine too, though, that while (let's assume) our use of the term 'good' is causally regulated by one kind of natural property (say, pleasure) theirs is regulated by a different natural property (say, being the object of a desired desire); and we might imagine too that while our use of 'right' is regulated by one kind of natural property (say, conduciveness to maximizing

over-all pleasure), theirs is regulated by a different natural property (say, conformity to certain socially accepted rules).<sup>20</sup>

The thing to notice, the Gracelanders point out, is that supposing our terms and theirs are causally regulated by different natural properties does nothing to undercut our sense that their terms mean what ours do, in light of the dramatic similarities between their practice and ours. We no doubt differ from them in what we count as good and right, but this difference reflects, the Gracelanders suggest, a disagreement not rightly characterized as a difference in meaning or reference.

This is just as it shouldn't be if our moral terms referred to whatever natural property causally regulated their use. For if that were what determined the reference of our moral terms then we should see their 'moral' terms as differing from ours in just the way we see Twin Earthers' use of 'water' as differing from ours. We ought to be comfortable in treating their claim that some course of action is right (on the grounds that it conforms to certain social conventions) as unproblematically compatible with our claim that it is wrong -- after all, in making these claims we are (supposedly) ascribing fundamentally different properties to that course of action. But this isn't how things go. Instead, assuming that Twin Earthers deploy their 'moral' terms in a practice much like ours, we treat their claims as having the same meaning they would have in our mouths despite the fact that what happens to causally regulate their use of the terms is different. We see ourselves as disagreeing with them, not talking past them.

Moreover, while Twin Earthers may well have a word, say 'pleasure', that does refer to pleasure, that term should not be translated by our 'good', if pleasure in Twin-USA doesn't play the action-guiding role in their community that it does in ours. As the Gracelanders see things, what matters, when it comes to translating evaluative terms, is not what natural property regulates the use of the terms but instead the roles those terms (or the properties they refer to, if they refer) play in governing behavior. The meaning of moral terms is best viewed, they suggest, not as being determined by the real nature of what causally regulates the use of the terms but as being a reflection of the terms' action-guiding role within a community.

This argument can be put as a revised version of the Open Question Argument:<sup>21</sup> Granted, the fact that it is an open question whether pleasure is

---

<sup>20</sup>. Of course, we needn't travel to Twin Earth for this sort of example, since it is a familiar fact of life that different communities rely on different criteria in applying their moral terms and as a result are responsive to different natural features in reaching their moral views.

<sup>21</sup>. Horgan and Timmons put the argument this way in "Troubles for New Wave Moral Semantics," p. 163.

good doesn't by itself show that 'good' and 'pleasure' refer to different properties. But the fact that it remains an open question, even given the assumption that pleasure is in fact what causally regulates our use of 'good', does show (the Gracelanders argue) that the terms do not refer to the same property. They put the argument this way: Competent speakers of English will treat as closed the question "Is H<sub>2</sub>O water?" given the assumption that H<sub>2</sub>O causally regulates the use of our word 'water.' Similarly, if the new semantics applied to moral terms, competent speakers of English would treat as closed the question "Is pleasure good?" given the assumption that pleasure causally regulates the use of our word 'good'. Yet, they don't. And this shows, the Gracelanders contend, that the new semantics, adequate as it might be to capturing the meaning and reference of names and natural kind terms, is wholly inadequate as an account of the semantics of moral terms.<sup>22</sup>

#### Moral Kind Terms

The Gracelanders' argument reflects an important fact, but I think they draw the wrong conclusion. The important fact is that we do not defer to science in settling whether the property that regulates other communities' use of some word (say 'good') is of *the same kind* as what regulates our use of 'good', for the same reason we do not defer to science in determining the reference of our term 'good'. The kinds that matter to morals are not those of natural science but those that are countenanced by the best moral theory. Whether the Twin Earthers' term 'good' refers to what our term does depends on whether their use of the term is causally regulated, in the appropriate way, by what is good, and it may well be so regulated even if all the instances of goodness with which they come into contact are quite different in other respects -- even if, for example, they are all instances of preference satisfaction but not pleasure.

---

<sup>22</sup>. This way of putting the Revised Open Question Argument is at least misleading. Defenders of the new semantics do not hold that competent speakers of English will treat the following questions (I quote from Horgan and Timmons) as closed: "Given that the use of 'water' by humans is causally regulated by the natural kind H<sub>2</sub>O, is liquid L, which is H<sub>2</sub>O, water?" and "Given that the use of 'water' by humans is causally regulated by the natural kind H<sub>2</sub>O, is liquid L, which is water, H<sub>2</sub>O?" Thinking *these* must be closed questions would involve holding that the positive proposals about how natural kind terms get their reference (by being causally regulated by the individuals or properties to which they refer) constitute truths recognizable by competent speakers simply in virtue of their competence. But that is not the status of these proposals. Instead, they represent empirical/philosophical hypotheses concerning the semantic properties of some of our words. Of course, if the proposals are correct then they must explain the judgments people make concerning the reference of their terms. They must account for how and why people come to think of the terms as referring to this or that individual or kind. So they must in this way respect the linguistic intuitions of competent speakers. Yet, for all that, respecting and explaining these intuitions is compatible with advancing an explanation the truth of which is not recognizable by competent speakers simply in virtue of their linguistic competence with the term itself.



The point here is not that people simply don't believe that pleasure constitutes the real (as opposed to nominal) nature of goodness, although I would guess most people don't. That is not the issue. The point is rather that, when it comes to our moral terms, discovering (or stipulating) that our use of the term happens to be causally regulated by things that all fall into the extension of some natural term (say, 'pleasure'), doesn't settle for us that 'good' rigidly designates *that* property. For an analogous reason, when it comes to the term 'water', discovering that our use of the term happens to be causally regulated by things that all fall into the extension of 'in liquid form at 60 degrees Fahrenheit' doesn't settle for us that 'water' rigidly designates *that* property.

If moral terms work as natural kind terms do, it will be by rigidly designating whatever bears the *relevant* same-kind relation to the samples of that kind that appropriately regulate our use of the terms. But the kinds that are relevant are different. In the case of natural kind terms, samples will bear the requisite relation if and only if they fall within the same natural kind, and it is irrelevant whether they are, in other respects, all of the same kind. Whereas, in the case of moral (kind) terms, samples will bear the requisite same-kind relation if and only if they fall within the same moral kind, and it is irrelevant whether they are, in other respects (say, from the point of view of science), all of the same kind. What we need to know about the Twin Earthers is whether their use of the term 'good' is appropriately regulated by what is in fact good. To learn only that it is regulated by things that fall into some other (natural) kind is not yet to learn enough.

The crucial Moral Twin Earth thought experiment is one in which we suppose that the Twin Earther's use of 'good' is causally regulated not by what is (in our view) good but by something else, just as in Putnam's examples we are to suppose that their use of 'water' is causally regulated not by what is (in our view) water but by something else. Of course, Putnam's examples work to reveal that, in our view, being composed of something other than H<sub>2</sub>O establishes that something is not water, but it is no part of the semantic theory that we do. Instead, the semantic theory explains why, given that we do hold this empirical view, we count the Twin Earther's 'water' as having a different meaning from ours -- the terms have different meanings because the terms' meanings are not determined by the concept or description we and they associate with our respective terms, but (in part) by the nature of the world in which they find themselves. It also explains why, were we to change that view, we would change as well our judgments as to what would count as a sample of water.

H<sub>2</sub>O gets into the picture, as the property rigidly designated by 'water', because we accept the view (recommended by the best scientific theory) that being composed of H<sub>2</sub>O molecules is the real nature of what we've been referring to all along. It is that commitment that leads us to hold that the Twin Earther's 'water' refers to water only if it is causally regulated in the appropriate way by samples of H<sub>2</sub>O. Similarly, pleasure (or preference satisfaction, or some

other natural property) will get into the picture, as the property rigidly designated by 'good', only if we accept the view (that might be recommended by the best normative theory) that pleasure (or whatever) is the real nature of what we've been referring to all along. It is that commitment, were we to have it, that would lead us to hold that the Twin Earther's 'good' refers to goodness only if it is causally regulated in the appropriate way by pleasure (or preference satisfaction or some other natural property).<sup>23</sup>

If the semantics of kind terms extends to moral terms, then what we should find is that people who accept hedonism as the right account of value will be willing to say that on Twin Earth, if the Twin Earthian use of the term 'good' is causally regulated in the appropriate way by something other than pleasure, then their term does not after all mean what ours does, regardless of what action guiding role it might play. Some confirmation of the semantic view is actually found, I think, in the tendency of people who reject the implication to reject hedonism as well. Insofar as one holds that something other than pleasure might be good, one will resist taking the fact that people on Moral Twin Earth refer to something other than pleasure as sufficient to establish that they are not referring to goodness -- even if one thinks that, in our world, every instance of goodness is pleasant and every instance of pleasure is good.

We should find too that people who accept the view that to be good is to be the object of a desired desire will be willing to say that on Twin Earth, if the Twin Earthian term is causally regulated in the appropriate way by something other than being an object of desired desires, then the Twin Earthian term does not after all mean what ours does, regardless of what action guiding role it might play. Again, some confirmation of the semantic view is found, I think, in the tendency of those who reject the implication to reject the account of value in terms of desired desires too. And again, insofar as one holds that something other than being the object of desired desires might be good, one will resist taking the fact that they refer to something other than the object of desired

---

<sup>23</sup>. The best scientific theory presumably reveals natural kinds, and that is why, in our use of natural kind terms, we fairly systematically defer to (what we take to be) the best science in identifying their referents. Thus we change our view about what (if anything) our natural kind terms refer to as we change our view of the kinds recognized by the best science. In our use of moral kind terms, however, we defer not to science but in effect to (what we take to be) the best normative theory. And we change our view about what (if anything) our moral kind terms refer to as we change our view of the kinds recognized by the best normative theory. Of course, in the case of science, the pattern of deference our use exhibits often takes the form of deferring to the judgments of scientists whereas, with moral terms, people generally defer much less to the views of others. However, this difference is, I think, a difference only in degree. It reflects peoples' differential willingness to recognize others as more expert than themselves, not a difference in the role played by what we take to be the best relevant theory -- where that role is the role of determining when various actual and possible things bear the relevant relation to one another of being-of-the-same-kind.

desires as sufficient to establish that they are not referring to goodness -- even if one thinks that, in our world, every instance of goodness is an object of a desired desire and every instance of an object of desired desire is good.

Indeed, what evidently does travel in tandem is, on the one hand, our willingness to see the Twin Earthian term 'good' as referring to what ours does and, on the other, our seeing their term as regulated by what is good (rather than some possible simulacrum); just as, in the case of 'water', what travels in tandem is, on the one hand, our willingness to see the Twin Earthian term as referring to what ours does and, on the other, our seeing their term as regulated by what is water (rather than some possible simulacrum). And it is the fact that these views travel in tandem in both cases that recommends thinking of the terms 'water' and 'good' as referring to the relevant kind that causally regulates their use rather than to whatever happens to satisfy some concept or description they or we associate with the term.

Needless to say, there is not the same consensus about the nature of value, as there is about the nature of water. And the lack of consensus makes working out the Twin Earth examples more difficult. In effect, the examples need to be tailored to the various views one might have concerning the nature of value, since the force of the examples depends on describing Twin Earth in a way that has people there using the terms in response to what superficially resembles, but is not actually the same as, what we take ourselves to be referring to. Still, worked out in the right way, I think such examples would offer substantial support for the view that our moral terms work to refer in very much the way our natural kind terms do. For in each case, as the examples play out against a background account of the nature of value, I suspect people will find they regard the Twin Earthers' 'good' as meaning the same as ours when, but only when, they see its use as appropriately responsive to value.

Of course we needn't go so far away as Twin Earth in order to get examples of the sort of thing the Gracelanders have in mind. Right here on Earth it's not hard to find people and even whole communities whose use of moral terms reflect a sensitivity to natural features of a situation that others pretty much disregard. The most striking cases are those where the people in question explicitly embrace different criteria for the application of their terms. Thus those won over by a Rawlsian theory of justice may count as just only those institutions that more or less conform to the two principles of justice Rawls identifies, while others of a more utilitarian bent may, in determining which institutions count as just, keep a more or less fixed eye on whether they embody rules the adoption of which promotes over-all happiness, and still others will focus on whether the institutions enforce an egalitarian distribution of goods. Despite the differences in overt criteria, and so the difference in the natural features to which their use of the term 'justice' is sensitive, we see all these people not as talking past one another but as holding competing views about justice.

Such cases do, pretty clearly, tell against seeing the meaning of 'justice' as being determined by the criteria people deploy in using the term. And, as the Gracelanders would have us conclude, they tell against seeing 'justice' as referring to the natural kind (assuming there is one) that happens to causally regulate a person's (or community's) use of the term.<sup>24</sup> But they do not tell against mobilizing the semantics of natural kind terms, suitably adjusted for application to moral kind terms; they do not tell against seeing 'justice' as referring to the relevant -- in this case moral -- kind (assuming there is one) that causally regulates a person's (or community's) use of the term.

In fact, a crucial assumption lying behind our willingness to see Rawlsians, utilitarians, and egalitarians as all talking about justice is our sense that the various criteria they each rely on have emerged in response to (as we see it) samples of a single moral kind -- justice.<sup>25</sup> Often enough the criteria are, quite clearly, embraced on the grounds that they articulate the true nature of what they all agree we've been talking about all along.<sup>26</sup> When one or another criterion is embraced on these grounds, this will reflect the conviction that the best normative theory will countenance seeing the criterion as specifying the true nature of justice (in a way that parallels our conviction that the best scientific theory will countenance seeing H<sub>2</sub>O as specifying the true nature of water). And even when the criteria are not deployed consciously on these grounds, we will see those who use the various criteria as nonetheless talking about what we are so long as we see their reliance on the criteria as their more or less successful way of picking out instances of what we are referring to.<sup>27</sup>

---

<sup>24</sup>. This is basically the version of the argument offered by Simon Blackburn in "Just Causes."

<sup>25</sup>. Or perhaps as a response to samples of its opposite, injustice.

<sup>26</sup>. They won't always, of course, be embraced on these grounds; one might rely on a particular criterion -- believing it is a reliable way of determining what is just -- without believing it gets at justice's true nature, in the same way we rely on certain superficial features of a stone in judging whether it is gold without believing that having that appearance constitutes anything like the true nature of gold.

<sup>27</sup>. Although we recognize that not all that glitters is gold, we might well see those who apply 'gold' to whatever glitters as using it to refer to gold, if we view their reliance on glittering as a more or less successful attempt to pick out instances of what is actually gold. Whether we will see their reliance on glittering in this way will depend, among other things, on whether we think they would shift their criteria on learning more about the nature of what they've identified as 'gold'. If they wouldn't, we would have substantial grounds for thinking that they use 'gold' not to refer to gold but only to, say, what glitters. A key piece of evidence, when it comes to determining whether some term is a natural kind term, is found in whether competent users of the term exhibit an appropriate deference, in their application of the term, to information concerning the scientifically significant nature of the stuff they've identified as using the term. Similarly, a key piece of evidence, when it comes to determining whether some term is a moral kind term, is found in

To see the importance of this assumption, imagine a community in which the term 'justice' has the action guiding role highlighted by the Gracelanders but is not appropriately regulated by (instances of) justice. This is easy enough if we imagine their use as strictly confined in application to things that conform to extent social rules, where neither the rules nor what conforms to them are even approximately just. In such a case, the evident insensitivity to justice of their use of 'justice' would naturally underwrite seeing them as talking not about justice at all but instead about, for instance, merely what is socially approved. Of course, we would still, almost surely, see ourselves as disagreeing with them about what should guide action, yet it would be a mistake to see our term 'justice' and their orthographically identical term 'justice' as equally deployable in a discussion of justice.

Incidentally, this point does not turn on the so-called 'thickness' of our concept of justice, on the fact that there are descriptive constraints on what we might recognize as an example of justice. For the point goes through if we substitute in talk of what is right or good (which carry no substantial descriptive constraints). If we discovered of a community that their use of the terms 'right' and 'good' were not appropriately regulated by what is right or good but instead by something else we would again have grounds for thinking that they were not using the terms to say of things what we say with ours -- even if their terms played a role in guiding their actions. And again we would presumably have a disagreement with them, yet it would be a mistake to see our terms 'right' and 'good' and their orthographically identical terms 'right' and 'good' as equally deployable in a discussion of what is right or good.<sup>28</sup>

Our willingness to see others as using terms that are suitably translated by our terms 'just' or 'right' or 'good' turns on our seeing their use as appropriately regulated by institutions, actions, and characters that are (in their context) actually just, right, or right. Our grounds for thinking that their use is appropriately regulated will presumably turn on a number of considerations. It needn't be, though, that those using the term generally or reliably apply their terms only to what are instances of the relevant kind. It may be that ideology, ignorance, and self-interest all conspire in various and familiar ways to (mis)shape peoples' substantive views about the nature of justice, say, and so influence in an unfortunate way what they see as just. Justice itself might still

---

whether competent users of the term exhibit an appropriate deference, in their application of the term, to information concerning the morally significant nature of the actions, institutions, or characters they've identified using the term.

<sup>28</sup>. How exactly we should then formulate the disagreement is of course tricky. If they are, as we might discover, not at all concerned with what is right and good, then the disagreement will not be about whether it is right or good to have that concern. It seems instead best characterized as a conflict in attitude; we think they should be concerned, we might say, and they don't give a damn.

be implicated in the explanation of their use (and misuse) of the terms in a way that would justify our seeing them as using the term to refer to the relevant moral kind despite their mistaken views about its nature or about what counts as instances of that kind. What matters is whether, in misapplying the term to the variety of things they do, they are nonetheless talking about that kind.

In using moral terms, we take ourselves to be referring to that in virtue of which things count as really valuable, right, virtuous, or just. Yet we do not take our current views on these matters as decisive and certainly not as settling the reference of our terms. This is evidenced by our willingness (if not readiness) to alter our views without thinking we've thereby abandoned thinking about what is good, right, virtuous, etc. And it is shown as well in our readiness to recognize that those with dramatically different views are nonetheless holding views about what is good, right, etc. To the extent some other community's use of their terms can likewise be seen as an attempt to refer to that in virtue of which things count as really valuable, right, virtuous, or just we will have grounds for using our terms to translate theirs even if they embrace (as we see it) wildly mistaken views about the real nature of what we and they are talking about.

As the moral kinds approach would have it, what a moral term refers to, if anything, is determined by whether, in light of the best moral theory, the use of that term can be seen as appropriately regulated by instances of a normatively significant kind. Our sincere deployment of the terms reflects, in turn, our conviction that we are using them to refer to what the best theory would reveal to be normatively significant kinds. But that conviction travels with a recognition that we ourselves may be mistaken in quite significant ways about the shape the best theory might take.

It is not a forgone conclusion that our moral terms will find a reflection, let alone a vindication, in the best normative theory. Bentham's famous contention that talk of rights is "nonsense upon stilts" expresses his perfectly intelligible conviction that talk of rights picks out no normatively significant kind. What makes his claim intelligible is that the best normative theory might not count rights as a normatively significant kind -- just as developments in science have shown that 'phlogiston' doesn't refer to an explanatorily significant kind.<sup>29</sup>

Needless to say, the lack of consensus concerning the nature of value, rightness, justice, virtue... will well and reasonably undermine ones confidence in any particular normative theory. It raises obvious problems, too, when it comes to thinking of normative theory in general as getting more or less

---

<sup>29</sup>. See Jeremy Bentham's *Anarchical Fallacies*, in *Works* (Edinburgh: W. Tait, 1843), ii. 501.

successfully at the nature of value, rightness or virtue. My own admittedly optimistic view is that these problems can be met, that we can, in effect, justify seeing normative theory as a reasonably attempt to get clear on what is normatively significant. And we can do this even if we should not put much confidence in the particular theories we are actually able to develop.

3

Even if my optimism is misplaced, however, the lack of consensus doesn't seriously challenge the semantic proposal that our moral terms refer to whatever normatively significant kinds (if any) appropriately regulate their use. Nor does it undermine the suggestion that, in using moral terms, we take ourselves to be talking about, referring to, and arguing concerning one and the same thing, when we advance our own (apparently competing) views. The very fact that we treat differing views as competitors suggests this interpretation. And a substantial advantage of the moral kinds approach is that it articulates what commitments go along with seeing ourselves as disagreeing with, rather than talking past, others who hold radically different views. The crucial commitment is that there is, as we might put it, a fact of the matter about which we disagree, a fact capturable (even if not yet, or ever, actually captured) by a normative theory that spells out the true nature of what we are talking about.

Whether these commitments are, in the end, defensible cannot be settled by the semantics but will instead turn on whether there is, at least in principle, a defensible normative theory.<sup>30</sup>

---

<sup>30</sup>. I've benefitted substantially from conversations with and comments from Louise Antony, Simon Blackburn, Terence Horgan, Jaegwon Kim, William Lycan, Jeremy Ofseyer, Philip Pettit, Michael Smith, and Ernest Sosa.

