University of Massachusetts Amherst

ScholarWorks@UMass Amherst

Doctoral Dissertations

Dissertations and Theses

Spring August 2014

Synthetic Reductionism in Moral Philosophy

Scott Hill

Recommended Citation

Hill, Scott, "Synthetic Reductionism in Moral Philosophy" (2014). *Doctoral Dissertations*. 94.
https://doi.org/10.7275/01zv-c884 https://scholarworks.umass.edu/dissertations_2/94

SYNTHETIC REDUCTIONISM IN MORAL PHILOSOPHY

A Dissertation Presented

by

SCOTT HILL

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2014

Philosophy

SYNTHETIC REDUCTIONISM IN MORAL PHILOSOPHY

A Dissertation Presented

by

SCOTT HILL

Approved as to style and content by:

Fred Feldman, Chair

Peter Graham, Member

Lynne Baker, Member

Andrew Dole, Member

Joseph Levine, Department Head
Philosophy

DEDICATION

To Byron and Earlene Hill.

ACKNOWLEGEMENTS

I thank my dissertation director Fred Feldman as well as my other committee members Lynne Baker, Andrew Dole, and Pete Graham. I also thank Louise Antony, Phillip Bricker, Maya Eddon, Ernesto Garcia, Vanessa De Harven, Hilary Kornblith, Gary Mathews, and Christopher Meacham. I further thank Elle Benjamin, Brandy Burfield, Donovan Cox, Justin Dealy, Ed Ferrier, Matt Gifford, Bob Gruber, Ali Jawin, Dennis Kavlakoglu, Jordan Kroll, Haoying Liu, Tricia Magalotti, Peter Marchetto, Josh Moulton, Luis Oliveira, Kristian Olsen, James Patten, Bailey Peterson, Ben Rancourt, Julie Rose, Kim Soland, Miles Tucker, Aaron Washington, and Hayley Webster. Finally, I thank my parents, Byron and Earlene Hill.

A version of Chapter 1 appeared previously as ‘An Adamsian Theory of Intrinsic Value’, *Ethical Theory and Moral Practice* (2011): 273-289.

ABSTRACT

SYNTHETIC REDUCTIONISM IN MORAL PHILOSOPHY

MAY 2014

SCOTT HILL, B.S. KANSAS STATE UNIVERSITY

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Fred Feldman

My dissertation consists of five independent papers linked by their connection to a larger meta-ethical story to which I am sympathetic. Each chapter contributes in some way to exploring that story. To begin with, I will briefly sketch the story and say something about where my sympathies lie. Then I will discuss each of the chapters of my dissertation and how they fit into that story. To finish things, I will identify some aspects of the story that are not explored in my dissertation but stand in need of further defense.

The story contains a metaphysical thesis and an epistemological thesis. The metaphysical thesis is synthetic ethical reductionism. According to this view, moral properties are identical to properties that can be expressed without using moral terms; however, no moral term has the same meaning as any non-moral term. The epistemological thesis is theistic intuitionism. According to this view, God knows which moral properties are identical to which properties expressible in non-moral vocabulary and He created humans with moral intuitions that reliably indicate which things have those properties in which cases.

In Chapter 1, I contribute to this story by exploring the details of how a theological version of the metaphysical thesis might go. I start by describing Robert

Adams' theory of intrinsic value. According to Adams, intrinsic value and similarity to God are identical properties. I discuss some of the ways in which Adams' conception of intrinsic value differs from the more traditional conception of Mill, Moore, Sidgwick, Chisholm, and others. I then consider a series of problems for Adams' theory and identify solutions to those problems.

In Chapter 2, I contribute to the story by exploring the details of how a naturalist version of the metaphysical thesis might go. I consider some objections to naturalist synthetic ethical reductionism due to Robert Adams, Alvin Plantinga, and Michael Rea. I then offer responses to those objections on behalf of the naturalist.

In Chapter 3, I contribute to the story by pressing it on one of its problems. G. E. Moore's famous open question argument is widely seen as refuting analytic forms of ethical reductionism. Synthetic reductionists like me, however, often claim that the open question argument poses no threat to our variant of reductionism. Against this, Horgan and Timmons have offered the moral twin earth argument. A natural way of interpreting their argument is as an extension and modification of Moore's open question argument. While the original open question argument is often seen as inapplicable to synthetic reductionism, the moral twin earth argument is taken to be a serious threat to synthetic reductionism. I consider responses to the moral twin earth argument by various authors sympathetic to synthetic reductionism. I then show that these responses are unsuccessful and that the moral twin earth argument remains a threat to my story.

In the last two chapters, I contribute to the story by pressing an alternative story on one of its problems. The competing story is Mooreanism. A common allegation against Mooreanism is that it cannot accommodate moral knowledge while synthetic reductionism allegedly can. The worry is that, according to Mooreanism, moral properties are causally inefficacious. Given this assumption, it is difficult to see how our moral beliefs could be anything other than, at best, accidentally true. And so moral knowledge, if Mooreanism is true, is impossible.

In Chapter 4, I consider Michael Huemer's discussion of this problem. Huemer argues that, first, Mooreanism and synthetic reductionism are epistemologically on a par. So if Mooreanism is epistemologically problematic, then synthetic reductionism is as well. Second, Huemer sketches a theory of a priori knowledge that is supposed to show that Mooreanism can account for moral knowledge. I argue that Huemer is mistaken on both counts. Mooreanism and synthetic reductionism really are not epistemologically on a par. And Huemer's theory of a priori knowledge is problematic.

In Chapter 5, I contribute to the story by considering George Bealer's theory of a priori knowledge. Bealer's theory is supposed to explain how a priori intuitions, including moral intuitions, are reliable. I argue that Bealer's theory is problematic. First, it is impossible for anyone, other than God, to satisfy the conditions Bealer's theory places on a priori knowledge. Second, Bealer's theory is supposed to show that intuition is a basic source of evidence. But instead, his theory implies that intuition is not a basic source of evidence.

Some features of the story I am interested in are left undefended or unexplored in my dissertation. A full statement and defense of my view would require picking up on some unfinished business. One thing that is still needed is further discussion of the moral twin earth argument. In Chapter 2, I defend the moral twin earth argument from a series of objections. However, I do not offer my own objections to the argument. A full defense of my view would include such objections.

Another area needing further development is the philosophy of language on which synthetic reductionism rests. In Chapters 2 and 4 I say a few, very sketchy, things about this. But much more needs to be said to successfully defend synthetic reductionism.

A third area needing further discussion is an explanation of how God knows which moral properties are identical to which properties expressible without moral vocabulary. Chapter 4 includes a partial sketch of how this might go. But much more needs to be said.

TABLE OF CONTENTS

CHAPTER 1

AN ADAMSIAN THEORY OF INTRINSIC VALUE

In this paper I develop a theological account of intrinsic value drawn from some passages in Robert Merrihew Adams' book *Finite and Infinite Goods*. First I situate his theory within the broader literature on intrinsic value, and draw attention to some of its revisionist features. Next I state the theory, raise some problems for it, and refine it in light of those problems. Then I illustrate how the refined theory works by showing that it has the resources to deal with some seemingly formidable objections.

1.1 What Sort of Intrinsic Value?

The concept of intrinsic value Adams is interested in is very different than that of authors such as Mill, Moore, Sidgwick, Chisholm, and others. For this reason, it is important to say something about what Adams takes the bearers of intrinsic value to be, in virtue of what those objects are bearers of intrinsic value, and what sorts of slogans he uses to express the target of his inquiry.

Persons: For Adams, intrinsic value is first and foremost a property of persons. God (1999, p. 14, 113) is intrinsically valuable to the highest degree. He is "the supreme standard of value." All persons (1999, p. 115-21) are intrinsically valuable to a very large degree merely in virtue of their being persons. In addition to personhood, there are many other properties which, when had by a person, increase the degree to which that person is intrinsically valuable: loving goodness (1999, p. 131-2), being allied with and being for goodness (1999, p. 28), being creative (1999, p. 31), being healthy (1999, p. 116), being sexual (1999, p. 116), being rational (1999, p. 117), having various emotional and social capacities (1999, p. 116), being able to tell a good joke (1999, p. 116), having friendships (1999, p. 119), being able to cook well (1999, p. 30), being athletic (1999, p. 83), being philosophical (1999, p. 83), caring about the right things and the effectiveness of that caring (31), perceiving a situation accurately (1999, p. 31), being beautiful (1999, p. 31), being virtuous (1999, p. 14).

Adams sometimes worries (1999, p. 117-8) that his concept of intrinsic value is unacceptably inegalitarian. He worries that since different persons have these properties to very different degrees and since some persons have many more of these properties than others, he will be forced to hold that some persons are more intrinsically valuable than others. It is unclear what Adams' final response to this worry is. In some cases he suggests (1999, p. 117) that there are differences in intrinsic value between human persons, but the degree to which each such person is valuable in virtue of being a person is much greater than the differences in degree of value between them. In other cases he suggests (1999, p.) that a very morally corrupt person is of much less intrinsic value than other persons. In still other cases he suggests (1999, p. 117-8) that there is no way to compare the value of one person to another or that there may be a way but we shouldn't or that there is no such comparison to be made. Again, it is unclear what Adams' final position on this is.

Living Organisms: While persons, for Adams, are the primary bearers of intrinsic value, they are not the only objects that have it. All living organisms have intrinsic value[1] merely in virtue of being living organisms. Examples of non-human organisms to which

[1] Adams (pp. 116-7) says that "I think the excellence that all of us possess… begins with the simplest functions of animal and vegetative life…. [I]t is particularly important for me to ascribe excellence to these unglamorous natural functions.... Surely it is a good for me (and not just an instrumental good) to enjoy good health, or simply to enjoy my physical, animal being, or living. I must therefore avoid a snobbish sense of 'excellent' in which that would not be enjoying something excellent."

Adams attributes intrinsic value include flowers (1999, pp. 28, 117), dogs (1999, p. 117), and sheep (1999, p. 117). As in the case of persons, there are further properties, such as being healthy (1999, p. 117), that when had by a living organism contribute to the degree to which it is intrinsically valuable.

Artifacts: The products of creative work are capable of having intrinsic value as well. For Adams, gourmet meals (1999, p. 30), paintings (1999, p. 83), poems (1999, p. 17) mathematical proofs (1999, pp. 17, 83)[2], novels (1999, p. 83), and performances (1999, p. 83) are among the creative achievements that are capable of having intrinsic value.

Nature: Natural phenomena, such as beautiful sunsets (1999, p. 83), are capable of having intrinsic value as well.

Worlds: Adams discusses (1999, pp. 17, 118-21) the application of his concept of intrinsic value to worlds. There are hints[3] that he does not think worlds are capable of having intrinsic value. But he seems to think the issue is debatable. Insofar as worlds are bearers of intrinsic value, Adams holds that the intrinsic value of a world is determined by the intrinsic values of the objects in that world. However, he does not think the value of a world is the simple sum of the intrinsic values of the objects in that world. For example, he thinks that a world with some people is better than a world with no people. But, even other things being equal, a world with more people is not always better than a world with fewer people. "Beyond a certain point" he says (1999, p. 119) "there is no

[2] Adams (p. 17) identifies poems and proofs as abstractions. Perhaps he thinks that some abstract objects can be bearers of intrinsic value.

[3] For example, he says (p. 118) "Though skeptical, myself, of global evaluations of worlds, I shall… argue that even if… worlds do have definite values, the most plausible assignments of value do not support the view that the value of persons… [implies] that the existence of each of them makes the world as a whole better."

advantage at all in sheer growth in numbers. Personhood is excellent, but it does not follow that more of it is better, in any way." These considerations extend to other, non-person, objects. To put it in the form of a slogan, Adams seems to endorse the following ideas about the contributions intrinsically valuable objects in a world make to the intrinsic value of that world: some is better than none, more variety is better than less variety, more of the same is not better than less of the same when there is already a lot of the same. Beyond these considerations, Adams says very little about the intrinsic value of worlds.

Properties: Adams also discusses the application of his theory of intrinsic value to properties. Like worlds, the intrinsic value of properties is not determined in the same way as the intrinsic value of objects. An intrinsically valuable property, for Adams (1999, p. 40), is a property in virtue of which an object that has that property has intrinsic value.

Adams uses (1999, p. 14) 'intrinsic value', 'goodness', and 'excellence' to refer to the property he is interested in. He attempts to express this concept in the following passages:

> What sort of good is at issue here? It is not *usefulness*, or merely instrumental goodness. It is not *well-being*, or what is good for a person. It is rather the goodness of that which is worthy of love or admiration. We have no word that in common usage signifies precisely and uniquely this kind of goodness; I shall refer to it often (though not always happily) as "excellence," and sometimes, (where I see on the horizon no confusion with other sorts of goodness) simply as "goodness" or the "good." Moral virtues are excellences in this sense, but Platonic excellence is not exclusively moral; beauty is a prime example of it…. These features of a Platonic structure for the realm of value are also features of the framework for ethics that I will present here (1999, pp. 13-4).
>
> There is *usefulness*, or merely instrumental goodness…. [There is] *well-being* or welfare…. The theory developed here, however, gives primary place to excellence…. It is the goodness of that which is worthy of love or admiration, honor or worship, rather than the good (for herself) that is possessed by one who is fortunate or happy, as such (though happiness may also be excellent, and worthy of admiration) (1999, p. 83).
>
> Like Plato's, mine will be a theory of nonmoral as well as moral value. The divine greatness adored in the Bible, by the mystics, and in the traditions of theistic worship is by no means

> exclusively moral. God is the exemplar or standard of a goodness that includes much more than moral virtue. Among the Platonic roles inherited by God, so to speak, is that of Beauty itself (1999, p. 14).
>
> The healthy functioning of living things is an integral part of much of the beauty of nature, and beauty of any sort is a prime example of what I mean by 'excellence' (1999, p. 116).
>
> Worship, after all, is the acknowledgement, not just of God's benefits to us, but of the supreme degree of intrinsic excellence (1999, p. 14).

In addition to what has been said already, there are three important respects in which the target of Adams' inquiry seems to differ from more traditional conceptions of intrinsic value:

Intrinsic Value is Not Intrinsic: Tradition holds that intrinsic value is an intrinsic property. Adams (1999, p. 14) holds that it is a relational property.

Intrinsic Badness is Derivative: Tradition holds that intrinsic goodness and intrinsic badness are both fundamental. Adams (1999, pp. 102-4) holds that only intrinsic goodness is fundamental. All sorts of badness, including intrinsic badness, are derivative and are to be analyzed in terms of fundamental intrinsic goodness.

Extrinsic Value Alters Intrinsic Value: Tradition holds that the extrinsic value of an object cannot alter that object's intrinsic value. Adams' theory (1999, p.) allows that the extrinsic value of an object can alter that object's intrinsic value.

1.2 Adams' Theory of Intrinsic Value

Adams (1999, p. 37) analyzes intrinsic value in terms of similarity to God. The simplest formulation of his theory is this:

GOOD: An object, *o*, is intrinsically good to degree, *n*, if and only if *o* is similar to God to degree *n*.

Adams (1999, p. 33) does not accept this simple formulation of the theory. There are examples that he thinks it gets wrong. Consider Hitler. Hitler was powerful. God is powerful. Hitler's power, therefore, contributed to the degree to which he was similar to

God. So if GOOD is true, then Hitler's power contributed to the degree to which he was intrinsically good. But Hitler's power did not contribute to the degree to which he was intrinsically good. So, GOOD is false.

In light of examples such as these Adams modifies the simple formulation of his theory. It is not just any sort of similarity to God that contributes to the degree to which an object is intrinsically good. It is only *faithful resemblance* that counts. Faithful resemblance, moreover, is "holistic." It requires more than the sharing of a single property. Adams' (1999, pp. 32-3) dense discussion of faithful resemblance is worth quoting at length:

> Probably the best reply we could give to these counterexamples [such as the Hitler example] if we want to defend the theory that excellence consists simply in resembling God is that not every sharing of a property constitutes a resemblance. Judgments of resemblance are more holistic than that…. The way or context in which a property is shared affects whether the sharing constitutes a resemblance or makes things more similar than they would otherwise have been…. There is a more pressing difficulty, however, for the thesis that Godlikeness is sufficient for excellence. More decisive counterexamples to it are possible involving holistic resemblance rather than mere sharing of properties. Consider the phenomenon of parody or caricature. Parodies and caricatures do resemble, but do not in general share the excellences of their original or object…. It is natural enough to say that Hitler's power is "a caricature of the divine power"—more natural, I suspect, than to deny flatly that his power resembles God's in any way. This suggests a modification of the analysis of excellence in terms of resembling or imaging God. We can suppose that the difference between resemblances to God that do and do not constitute virtues or excellences is analogous to the difference between good portraits (by which I mean faithful portraits) and caricatures. The excellence of other things besides God will consist, then, in the *faithfulness* of their imaging of God. This cannot be more than an analogy, of course, due to the radical imperfection of any resemblance of creatures to God. I will not offer here a full account of what the faithfulness of a portrait amounts to, and I am not sure that I could give one. It would surely include the observation that caricatures are *distorted* in a way that faithful portraits are not. The caricature exaggerates one or more features of the original, whereas the faithful portrait represents features in a balanced way and in relation to those other features to which they are most importantly related in the original. How much, and what, must be included in a faithful image depends on what is most *important* about the way in which the original has the features shared or represented.

Here Adams makes two revisions to GOOD. First, GOOD employs an account of similarity according to which any property that an object shares with God contributes to the degree to which that object is similar to God. But, in the present passage, Adams claims that similarity is “holistic.” Merely sharing a single property, he says, is insufficient for similarity.

Second, Adams (1999, pp. 34-6) says that among the ways of being similar to God in this more holistic sense, only faithful resemblance to God is what counts. An object that faithfully resembles God is like a portrait. An object that is similar to God, but does not faithfully resemble Him, is like a caricature. Although caricatures share properties with their original, they do so in a way that is “distorted” and “exaggerated” and “unbalanced”. The same is true of objects that do not faithfully resemble God. Furthermore, whether shared properties are distorted or exaggerated or unbalanced depends on what is “important” about the way in which God has those properties. As the degree to which an object has the relevant properties in an undistorted, unexaggerated, and balanced way increases, the degree to which that object faithfully resembles God also increases. The modified formulation of Adams’ theory, then, is this:

GOOD*: An object, *o*, is *intrinsically good* to degree, *n*, if and only if *o* faithfully resembles God to degree *n*.

Adams introduces the requirement about faithfulness in order to deal with the Hitler Objection. Perhaps this is the response he has in mind: Hitler was powerful. God is powerful. But merely sharing properties with God does not contribute to similarity to Him. Perhaps, in the relevant holistic sense, Hitler was similar to God. But Hitler was a caricature of God. He shared some properties with God, such as powerfulness; but the way in which Hitler had those properties was distorted, exaggerated, and unbalanced. So, GOOD* delivers the judgment that Adams wants—in spite of his great power, Hitler was not intrinsically good to a very large degree.

Adams’ reply to the Hitler Objection seems plausible when the judgment GOOD* delivers about Hitler is contrasted with the judgments it delivers about other subjects.

Jones: Jones is a very loving person. He is dedicated to the poor and spends a lot of time and effort trying to help them. Unfortunately, though, Jones is almost completely powerless. In spite of this, Jones' wealth of love motivates him to go out of his way to help people at great expense to himself. For example, he is not a competent speaker and he suffers from intense social anxiety. But, hoping to convince at least one person, Jones spends many sleepless and miserable nights working on speeches so that at public forums he can argue on behalf starving children. He hopes that they will be given food and shelter and aid in becoming more similar to God. He hardly ever succeeds. There are other examples of this: Jones barely makes enough money to get by. But he often goes without eating because he has given the change that he would have used for lunch to Religious Charity. Unbeknownst to him, however, the money is often misused. It rarely gets to the starving children for whom Jones intended it. While his lack of power prevents his efforts from being very effective, Jones helps people in whatever little ways that he can.

Smith: Smith is enormously powerful and a little loving. While Smith does a lot of good things for people, he would not bother if doing so were any more difficult for him. For example, Smith is a talented speaker. If he is at a coffee shop and people are having a discussion about whether or not it is important to help starving children, Smith will often step in and, in a matter of seconds, effortlessly convince them to give generously. He is always successful and people are changed forever after these brief coffee shop encounters. If it were any harder to do, however, Smith would have no hesitation about stopping. There are other examples of this: Smith makes a lot of money.

If a Religious Charity representative comes to Smith's door asking for a donation, Smith writes them a big check just to get them out of his hair so that he can get back to his main passion—listening to electronic music and drinking beer. As it so happens, Smith is by far Religious Charity's biggest donor. The money is always used effectively—securing for starving children food and shelter and aid in becoming more similar to God. While he is slightly pleased by this, Smith is mostly just happy that he can continue to wallow in his electronic music and beer drinking.

Max: Max is very loving and very powerful. He makes a lot of money. He gives most of it to the poor. Although he very much enjoys drinking beer and listening to electronic music, he would never think of spending his money or time on those things. He always devotes his time and money to the poor.

Min: Min is only a little loving and only a little powerful. He makes hardly any money. He never gives any of it away. He squanders what little he has on beer and electronic music.

GOOD*, when supplemented with the right way of construing distortion, exaggeration, and unbalance, seems to deliver correct, or at least acceptable, judgments about the comparative intrinsic values of Hitler, Jones, Smith, Max, and Min. Consider Jones and Smith. They are both similar to God in the relevant "holistic" sense. To varying extents they are both powerful and loving and so on. But their power and love are distorted and exaggerated and unbalanced. Of course, the way in which Hitler has the relevant properties is much more distorted and exaggerated and unbalanced than the ways in which Jones and Smith have the relevant properties. So Jones and Smith both faithfully resemble God to a greater degree than Hitler and are therefore intrinsically valuable to a greater degree than Hitler. Now contrast Jones and Smith with Max. Max has the relevant properties in a way that is not distorted or exaggerated or unbalanced at all. What is more, he has the relevant properties in greater proportions than Jones and Smith. So Max faithfully resembles God to a greater degree than Jones and Smith. So he is intrinsically valuable to a greater degree than Jones and Smith. Now contrast Jones and Smith with Min. Min has the relevant properties in a way that is much more balanced than Jones and Smith. But Jones and Smith have at least one of the relevant properties to a much greater degree than Min. The increase that each has in resemblance outweighs the decrease in lack of balance. So Jones and Smith each faithfully resembles God to a greater degree than Min. So Jones and Smith are intrinsically valuable to a greater degree than Min. Finally contrast Jones and Smith. Depending on which account of distortion and exaggeration and unbalance it is combined with, GOOD* will deliver a different judgment about the comparative intrinsic values of Jones and Smith.

1.3 Against Faithfulness

One odd feature of *Finite and Infinite Goods* is that Adams never offers an account of overall intrinsic value. He (1999, p. 37) offers an account of intrinsic goodness. He offers (1999, p. 103-4) an account of various sorts of badness and makes some suggestive comments about intrinsic badness. But he never unifies them in a way that generates an account of overall intrinsic value. Furthermore, it seems to me that the intuitions motivating Adams' revisions to GOOD are really intuitions about Hitler's overall intrinsic value rather than Hitler's intrinsic goodness. As I see it, the intuition is not that Hitler's power did not contribute to the degree to which he was intrinsically good. Rather, the intuition is that Hitler was so intrinsically bad that whatever intrinsic goodness he had is vastly outweighed by his intrinsic badness. To bring this intuition out, note that GOOD* says nothing about overall intrinsic value. Depending on just how much Hitler's power, together with his other properties, being distorted and exaggerated and unbalanced is taken to detract from his faithful resemblance to God, GOOD* either delivers the judgment that Hitler was intrinsically good to some positive degree or it delivers the judgment that Hitler was intrinsically good to degree zero. So unless the theory is unified with an account of intrinsic badness, Hitler was either overall intrinsically valuable to some positive degree or he was overall intrinsically valuable to degree zero or the theory delivers no judgment about Hitler's overall intrinsic value at all. To get Adams' theory to deliver the judgment that Hitler was overall intrinsically bad, therefore, an account that unifies Adams' discussion of intrinsic goodness and intrinsic badness is still needed. As I will argue below, there is a simple way to combine GOOD with Adams' theory of badness that delivers the judgment that Hitler was overall intrinsically bad. Since the unified account of overall intrinsic value is needed anyway and since the unified account, when combined with GOOD, can deliver the desired judgment, the move to GOOD* and to faithful resemblance is unmotivated.

Another worry I have about the move to GOOD* is that the crucial notions of distortion and exaggeration and unbalance seem to be very obscure. Adams says that what counts as distortion or exaggeration or unbalance depends on which properties are important and what is most important about the way in which God has those properties. But without further elaboration, importance is just as obscure as the notions it is used to explain. Fortunately, Adams (1999, pp. 33-4) does elaborate on importance. But I do not think the way in which he develops his account of importance is very promising. He says this:

> Our discussion has turned up important problems about resemblance—problems in determining whether the relationship between two things constitutes a resemblance and whether an acknowledged resemblance images God "faithfully" enough to constitute an excellence. It appears that both questions may turn on how *important* the shared properties in the case are…. For a theistic theory it is natural to appeal to God's view of things or God's attitude toward things. The theist can appeal to God's view of things as the definitive standard of importance, and hence of resemblance and of faithfulness, at least where the question is about images of God—saying that considerations in these matters have the importance God sees them as having, and that they resemble, and faithfully image God, insofar as God sees them as doing so.

A number of concerns spring up at this point. One might worry that such an appeal is circular. Indeed, Adams (1999, p. 34-8) spends a lot of effort arguing that it is not. Or, one might worry that which properties God sees as important, and which ways in which God has his properties are seen by Him as important, could have been very different than they actually are. So although, given what God actually sees as important, Hitler's power and other properties are had in a way that is distorted and exaggerated and unbalanced, if God had seen importance differently, Hitler's power and other properties would not be distorted or exaggerated or unbalanced in the relevant sense. So GOOD*'s judgment about Hitler would match GOOD's judgment about Hitler. So, if Adams' reasons for disapproving of GOOD are sound, then GOOD* is hardly an improvement. Of course, there are maneuvers Adams can make to try to avoid this problem. Maybe it is an essential property of God to see importance in the way that He does. Maybe Adams can show that positing this essential property does not lead to circularity. My only point is that all of these complications can be avoided by modifying Adams' theory in the way that I suggest below. My version of Adams' theory is compatible with, but does not need to posit, such an essential property. That is a reason to prefer my version of Adams' theory over GOOD*.

1.4 The Refined Theory

Distinguish between fundamental intrinsic goodness and overall intrinsic value. Rather than analyzing intrinsic goodness in terms of faithful resemblance to God, the refined theory returns to the original account of intrinsic goodness offered by Adams. I will call this sort of goodness "fundamental intrinsic goodness":

GOOD: An object, *o*, is *fundamentally intrinsically good* to degree, *n*, if and only if *o* is similar to God to degree *n*.

In *Finite and Infinite Goods*, Adams (1999, p. 103-4) discusses several kinds of badness and analyzes each in terms of similarity to God[4]. It is possible to analyze overall intrinsic value in terms of both fundamental intrinsic goodness and these kinds of badness.

The refined theory includes an account of instrumental badness (Adams 1999, p. 103). Consider some object. There may be objects that the object in question has caused to be less fundamentally intrinsically good than they once were. If there are, then for each of these objects, consider the degree to which its fundamental intrinsic goodness was decreased due to the pernicious influence of the relevant object. Then consider the sum of all such numbers. The resulting number is the degree to which the object in question is instrumentally bad:

BAD_I: An object, *o*, is *instrumentally bad* to degree, *n*, if and only if *n* is the number that is obtained by taking every object that is such that *o* causes a decrease in the degree to which that object is fundamentally intrinsically good and then adding each of those objects' decreases.

The refined theory includes an account of motivational badness (Adams 1999, p. 104). Motivational badness is to be analyzed in terms of a primitive againstness relation and fundamental intrinsic goodness. A person can be against an object. A person can be

[4] The accounts of instrumental badness and motivational badness are not necessarily meant to be accounts of our ordinary concepts of instrumental and motivational badness or anything like that. They are instead technical terms introduced to enable the theory to avoid the three objections that concern me in this paper.

against a property. The againstness relation can be both explicit and conscious or implicit and unconscious. If a person is against fundamental intrinsic goodness, then that person is motivationally bad. One way to be against fundamental intrinsic goodness is to be against some fundamentally intrinsically good object in virtue of that object's fundamental intrinsic goodness. Another way is to simply be against the property of being fundamentally intrinsically good:

BAD_M: A subject, *s*, is *motivationally bad* if and only if *s* is against fundamental intrinsic goodness.

The refined theory includes an account of intrinsic badness. Intrinsic badness is analyzed in terms of instrumental badness and motivational badness:

BAD: A subject, *s*, is *intrinsically bad* to degree, *n*, if and only if (i) *s* is motivationally bad and (ii) *s* is instrumentally bad to degree *n*.

The refined theory includes an account of overall intrinsic value. Consider the degree to which some object is fundamentally intrinsically good. Consider the degree to which that object is intrinsically bad. Subtract the later number from the former number. The resulting number is the degree to which that object is overall intrinsically valuable:

$GOOD_O$: An object, *o*, is *overall intrinsically valuable* to degree, *n*, if and only if *n* is the number that is obtained by subtracting the degree to which *o* is intrinsically bad from the degree to which *o* is fundamentally intrinsically good.

1.5 Return to the Hitler Objection

The theory will require further refinements. But, for expository purposes, it is useful to note that the theory, presently refined, delivers the correct judgment about the Hitler objection.

1.5.1 The "Hitler was Bad Overall" Reply

The first thing to do, in giving a proper response to the Hitler objection, is to change the desideratum. Adams seems to want the judgment that Hitler's power did not contribute to the degree to which he was fundamentally intrinsically good. But neither Adams' theory nor the refined theory can deliver that judgment. As I argued earlier, however, that is not the interesting issue. It is more important to make sure that GOOD_{O} delivers the judgment that Hitler was overall intrinsically bad to a very large degree. This can be done without denying that Hitler's power contributed to the degree to which he was fundamentally intrinsically good.

Here is a way that one might try to show that Hitler was overall intrinsically bad to a large degree: There were many people that Hitler hated who were fundamentally intrinsically good in various respects. So, by BAD_{M}, he was motivationally bad. Furthermore, Hitler killed a lot of these people and caused a decrease in the degree to which many of them were similar to God[5]. So, by BAD_{I}, Hitler was instrumentally bad to a very large degree. So, by BAD, Hitler was intrinsically bad to a very large degree. Now, Hitler was similar to God with respect to some properties. He was very powerful, he was smart, and he was a person. So, by GOOD, powerfulness, smartness, and personhood contributed to the degree to which he was fundamentally intrinsically valuable. But however much these properties contributed to the degree to which he was fundamentally intrinsically valuable, surely, whatever that turns out to be, it will be much smaller than the degree to which Hitler was intrinsically bad. Consider, therefore, the number that is obtained by subtracting the degree to which Hitler was intrinsically bad

[5] I should illustrate this point with some examples: Hitler subjected some Jews to experiments and withheld from them proper nutrition. These practices damaged their cognitive faculties and abilities. That is a decrease in the degree to which they are similar to God. Before the Holocaust there were many Jews who were members of communities. Hitler broke many of these communities up. Since God is a member of a communities (the Trinity and His people) that is a decrease in the degree to which they were similar to God. Hitler caused many Germans to hate Jews. Since God loves Jews that is a decrease in the degree to which those Germans were similar to God.

from the degree to which Hitler was fundamentally intrinsically good. It is a large negative number. So $GOOD_O$ correctly judges that Hitler was overall intrinsically bad to a large degree.

Although the refined theory is able to deal with this formulation of the Hitler objection, the present refinements are still not adequate. The problem reemerges given a slightly modified statement of the relevant example: There is some number that is the degree to which Hitler was intrinsically bad. Let it be any large number. Now, suppose there is someone, call him $Hitler_1$, who is like Hitler in the following respect: He does only and exactly the bad things that Hitler did. So he is instrumentally bad to the same degree as Hitler. What is more, there are a lot of people that $Hitler_1$ desires to kill and to make less similar to God. So he is motivationally bad as well. So, given BAD, $Hitler_1$ is intrinsically bad to the same degree as Hitler. Next, suppose that $Hitler_1$ is different from Hitler in the following respect: He is much more powerful. Since God is omnipotent, as a person's power increases, the degree to which that person is similar to God increases along with it[6]. So no matter how intrinsically bad $Hitler_1$ is, if he is stipulated to be powerful enough, he will resemble God and, hence, be fundamentally intrinsically good to a much greater degree. So suppose that $Hitler_1$ is powerful enough for this to be the case. Then consider the number that is obtained by subtracting the degree to which $Hitler_1$ is fundamentally intrinsically good from the degree to which he is intrinsically bad. The result is some large positive number. So, if $GOOD_O$ is true, then $Hitler_1$ is overall intrinsically good. But $Hitler_1$ is not overall intrinsically good. So $GOOD_O$ is false.

Here is another problem for the refined theory: Consider $Hitler_2$. $Hitler_2$ is

[6] This premise is questionable. But grant it for the sake of argument.

identical to Hitler with the exception that Hitler_2 is much less powerful than Hitler and any differences entailed by this exception. However, Hitler_2 caught many more lucky breaks than Hitler. So, in spite of his comparatively little power, Hitler_2 is instrumentally bad to the same degree as Hitler. Now contrast the overall intrinsic value of Hitler with the overall intrinsic value of Hitler_2. Other than the difference in power, they are identical in all respects relevant to the determination of their fundamental intrinsic goodness. So Hitler is fundamentally intrinsically good to a significantly greater degree than Hitler_2. Furthermore, Hitler and Hitler_2 are both motivationally bad and they are instrumentally bad to the same degree. So, by BAD, they are intrinsically bad to the same degree. Now subtract the degree to which Hitler is intrinsically bad from the degree to which he is fundamentally intrinsically good. Do the same for Hitler_2. Then contrast the two numbers. The number associated with Hitler will be significantly greater than the number associated with Hitler_2. So, if GOOD_O is true, then Hitler is overall significantly less intrinsically bad than Hitler_2. But Hitler is not overall significantly less intrinsically bad than Hitler_2. So GOOD_O is false. The refinements I have made to Adams' theory are not yet adequate.

1.5.2 The Fallen Properties Reply

In light of the above arguments, further modifications are needed. Fortunately, such refinements are available. The further modified theory includes an account of fallen properties:

FALLEN: A property, *p*, is *fallen* with respect to a subject, *s*, if and only if (i) *s* is motivationally bad, (ii) *s* is instrumentally bad to degree *n*, (iii) *p* is instantiated in *s*, (iv) *p* contributes to the degree to which *s* is similar to God and (v) *s*'s instrumental badness depends, in the relevant sense, on *s*'s instantiating *p*.

Contingent upon which dependence relation is taken to be relevant, condition (v) will be equivalent to one of these:

(v_1) if *s* had not instantiated *p*, then the degree to which *s* is instrumentally bad would have been less than *n*.

(v_2) *s*'s instantiating *p* is causally relevant to *s*'s being instrumentally bad.

(v_3) *p* is one of the properties on which *s*'s instrumental badness supervenes[7].

(v_4) *s*'s instantiating *p* grounds *s*'s being instrumentally bad[8].

There is no need, at present, to decide which of these dependence relations between the subject's instrumental badness and the relevant properties is the right one to employ. For present purposes I will interpret (v) as (v_1) and take the relevant sense of dependence to be counterfactual. Roughly and loosely put, I want to add the following idea to Adams' account of badness: it is bad to do bad things with a property that is good[9]. More

[7] Thanks to Fred Feldman for suggesting (v_3).

[8] Grounding is the relation discussed in Rosen (forthcoming) and Schaffer (2009).

[9] I have heard two complaints about modifying Adams' theory in this way. First, I have been told that the modification I introduce is "ad hoc" and that I am "gerrymandering" Adams' theory. Perhaps what someone who offers this objection means is that I introduce the fallen properties revision without any motivation other than the fact that it makes Adams' theory produce the correct judgment about Hitler and $Hitler_1$ and $Hitler_2$. But, such people seem to be claiming, modifications to a theory need independent motivation. So the fallen properties reply is defective.

Second, I have been told that I rely on a confusion about the relation between intrinsic and extrinsic value. My account of fallen properties implies that the intrinsic value of an object is sometimes altered by the extrinsic value of that object. But, as Moore has taught us, such a suggestion is incoherent and unintelligible. So my modification of Adams' theory is defective.

It is important to see that these objections are mistaken. Concerning the complaint about ad-hoc-ness and gerrymandering: The objector is holding me to an unreasonably high standard. Consider Feldman (1983). In some passages he introduces modifications to utilitarianism with no other motivation than the fact that those modifications make utilitarianism deliver the correct judgment about a class of cases. Sider (1991) and (1993) does the same thing. This is a common strategy. So the way in which I develop Adams' theory is in accordance with tradition.

I will, however, leave all of this to the side. I can meet the objector's unreasonably high standards. There is a precedent, in the literature, for the sort of move I have made. Shelly Kagan (1998, p.281) has argued, on the basis of several examples, that the extrinsic value of an object can alter its intrinsic value. He says this: "I want to leave open the possibility that the *intrinsic* value of an object may be based (in part) on its *instrumental* value." Much more recently, while commenting on Adams' theory, Kagan (2009a, p. 390) repeats this line. Josh Parsons (manuscript, p. 1) takes a similar line. He says this: "Suffice it to say that… objects can sometimes be instrumentally intrinsically valuable, and sometimes be finally extrinsically valuable… and let it be taken that people sometimes use "intrinsically valuable" when they clearly mean "valuable as-an-end", and use that misleading phrase to conflate the two." So there is a precedent in the literature, based on counterexamples to orthodoxy, for the sort of move I have made.

precisely, the modification I want to make is this: Consider some person who is both motivationally and instrumentally bad. That person may have some fallen properties. If that person does, then add the contributions that all of those properties make to the degree to which that person is similar to God. Then add the number just obtained to the degree to which the person in question is instrumentally bad. The resulting number is the degree to which that person is intrinsically bad. The modified account of badness, then, is this:

BAD$_2$: A subject, *s*, is *intrinsically bad* to degree, *n*, if and only if (i) *s* is motivationally bad and (ii) *n* is the number that is obtained by adding the degree to which *s* is instrumentally bad to the contributions made to the degree to which *s* is similar to God by all of *s*'s fallen properties.

The theory, when refined in this way, is able to deal with the variant Hitler objections. Consider all of Hitler$_1$'s fallen properties. Prominent among these is being powerful. For if Hitler$_1$ were not powerful, then he would not have been so instrumentally bad. Not included among Hitler$_1$'s fallen properties is having a weird moustache. So however much being powerful contributes to the degree to which Hitler$_1$ is fundamentally intrinsically good, by BAD$_2$, the same amount is added to the degree to which he is intrinsically bad. So no matter how powerful he is, that power, when all is said and done,

Therefore, although I do not need it, there is independent motivation for my introduction of fallen properties.

Concerning the second, Moorean, complaint: Bradley (2002) has argued persuasively that Kagan's examples can be accommodated by the Moorean. Bradley deals with the examples by saying that it is not objects but states of affairs that are the bearers of intrinsic value. With respect to my purposes, however, Bradley's response to Kagan is concessive. He allows that a state of affairs ascribing extrinsic value to an object can have intrinsic value. Here is what Bradley (2002, p. 26) says:

> Mooreans do indeed deny that for any x, the uniqueness or usefulness of x is relevant to x's intrinsic value. But I know of no evidence that Mooreans would deny that for any x, states of affairs ascribing uniqueness or usefulness to x could have intrinsic value. In fact, it must be pointed out that many philosophers within the Moorean tradition have thought that states of affairs ascribing non-intrinsic properties have intrinsic value.

In light of Bradley's work on this topic, the fallen properties reply, could easily be translated into terms that a Moorean would find unobjectionable. Following Bradley's strategy, one could recast Adams' theory in terms of states of affairs rather than objects. So a state of affairs ascribing fallen properties to an object would have less intrinsic value than one that did not. The relevant issue, therefore, is whether objects or states of affairs are the primary bearers of intrinsic value. Adams' theory can be formulated so as to accommodate either outcome.

does not make a positive contribution to Hitler_1's overall intrinsic value. What is left to do is to subtract Hitler_1's instrumental badness from the contributions made to the degree to which he is similar to God by his non-fallen properties. The result is some big negative number. Modifying the account of intrinsic badness in this way, therefore, enables GOOD_O to generate the correct result—Hitler_1 is overall intrinsically bad to a very large degree.

Now contrast once more Hitler and Hitler_2. By the reasoning given in the previous case, neither Hitler's power nor Hitler_2's power makes a positive contribution to their respective degrees of overall intrinsic value. Furthermore, the difference in power is the only difference between Hitler and Hitler_2 that is relevant to determining their intrinsic values. So the refined theory enables GOOD_O to deliver the correct judgment—Hitler is not significantly less intrinsically bad overall than Hitler_2. Neither the Hitler objection nor any of its variants pose a threat to the refined theory.

1.6 The "What if God Were Different?" Objection

Suppose that God is very different than He actually is with respect to His desires and His behavior[10]. Suppose that God hates academics. Suppose that He kills some of them and makes others of them stupid. Suppose that I also hate academics. Suppose that I also kill some academics and make others stupid. These properties contribute to the degree to which I am similar to God. So, by GOOD, they contribute to the degree to which I am fundamentally intrinsically good. So, given GOOD_O, these properties make a positive contribution to my overall intrinsic value. But killing academics and making them stupid does not contribute to my overall intrinsic value. So GOOD_O is false.

This formulation of the objection is not persuasive. The premise 'If GOOD_O is

[10] Adams (1999, p. 46) discusses this objection.

true, then these properties make a positive contribution to my overall intrinsic value' is false. To see this reason as follows: Academics are persons and they are smart. Even though God hates them, they are similar to Him with respect to these properties. So these properties contribute to the degree to which they are fundamentally intrinsically good. So, by BAD_M, I am motivationally bad. Note also that by destroying some academics and making the others stupid, I am causing a decrease in the degree to which they are similar to God. So given, GOOD, I am causing a decrease in the degree to which they are fundamentally intrinsically good. So, by BAD_I, these properties contribute to the degree to which I am instrumentally bad.

Now, note that my hatred of academics and my being such that I kill and render stupid academics are each fallen properties. After all, if I had not killed some of them and caused others of them to be stupid, then I would not have been so instrumentally bad. Furthermore, if I had not hated academics, then I would not have killed them and caused them to be stupid. So I would not have been so instrumentally bad. Given FALLEN, then, the relevant properties are fallen properties. So, since they are fallen, by BAD_2, the contribution the relevant properties make to the degree to which I am similar to God entails no corresponding contribution to the degree to which I am intrinsically bad. Furthermore, by BAD_2, the contribution that the relevant properties make to the degree to which I am instrumentally bad is added to the degree to which I am intrinsically bad. So the increase in similarity to God contributes nothing to my overall intrinsic value and the instrumental and motivational badness make a negative contribution to my overall intrinsic value. So, $GOOD_O$ delivers the correct judgment: if God hated and killed and stupidified academics and if I were to do those things, then such things would result in a decrease in the degree to which I am overall intrinsically valuable.

So far only one class of the relevant "What if God Were Different?" cases has been considered. What is common to each member of this class is this: There is some property that contributes to the degree to which various subjects are similar to God. God hates those subjects. God causes a decrease in the degree to which those subjects are similar to Him with respect to those properties. I am similar to God with respect to hating those subjects and causing a decrease in the degree to which they are similar to Him. In cases like these, the refined theory delivers the correct result. However, there is another relevant class of cases that cannot be handled so easily. There are "What if God were Different?" cases in which God tortures some subject but does not in any way cause a decrease in the degree to which they are similar to Him. The refined theory, one might reasonably think, delivers the wrong judgment about such cases.

Suppose that God hates academics. Suppose that once a week, just for fun, God causes academics to experience intense, horrific pain. Suppose, in addition, that God never causes a decrease in the degree to which these academics are similar to Him. If memories of such pain would cause the academics to neglect their studies and become less intelligent, then after each episode of pain God erases the memory. For any way in which such episodes of torture might result in a decrease in similarity to God, He takes steps to ensure that no such decrease occurs. So after each episode of pain, the academics' degrees of similarity to God are left constant. Before each episode of pain, the academics are smart. After the episodes, they are just as smart. Before, each episode each academic is a member of a community. After, they are still members of the same communities.

Now, suppose that I am very different than I actually am. Suppose that I share God's hatred of academics. Suppose that I find ways to cause them intense pain without producing any decrease in the degree to which they are similar to God. Maybe I build a machine that will allow me to do this. Maybe I am granted powers that will enable me to do this.

The refined theory, it would seem, delivers the wrong judgment about this case. It judges that, not only do the relevant properties contribute to the degree to which I am fundamentally intrinsically good, but they also make a positive contribution to my overall intrinsic value. To see this, reason as follows: My hatred of academics and my acts of causing them intense pain contribute to the degree to which I am similar to God. So, the relevant properties contribute to the degree to which I am fundamentally intrinsically good. Given BAD_{I}, and given that I never cause a decrease in the degree to which any academic is fundamentally intrinsically good, my being such that I perform these actions does not contribute to the degree to which I am instrumentally bad. Since the relevant properties do not contribute to the degree to which I am instrumentally bad, condition (v_1) of FALLEN is unsatisfied. So my hatred and acts of torture are not fallen properties. So, by BAD_2, such properties do not contribute in any way to the degree to which I am intrinsically bad. So, given GOOD_{O}, the relevant properties make a positive contribution to my overall intrinsic value. This is the wrong judgment.

Two replies are available on behalf of the refined theory. First, the premise of the above argument that reads 'my being such that I hate academics and torture them does not contribute to the degree to which I am instrumentally bad' is false. Consider again the motivation for this premise. The motivation was that any acts of torture would be erased from the memories of those who were tortured. So no psychological harm that could constitute or counterfactually determine a decrease in the academic's degree of similarity to God would obtain.

This motivation is inadequate. The academics in question do become less similar to God. Consider one of the academics the day after he was tortured. He reflects on the day before. He considers the proposition 'I was not tortured yesterday'. He believes that proposition. But his belief is false. So whereas before the academic had *n* false beliefs, now he has (at least) *n* + 1 false beliefs. So the academic is further away from having zero false beliefs than he was before. So since God has zero false beliefs and the academic has at least one more false belief than he did a few days ago, the academic is less similar to God than he used to be. And, of course, one does not have to actively consider a proposition in order to believe it. This result generalizes, therefore, to the other tortured academics who never actively consider the relevant proposition. So if I had not hated those academics, then I would not have tortured them. And if I had not tortured them, then they would not have the extra false beliefs. So if I had been without these properties, then I would not have been as instrumentally bad. So, by FALLEN, my being such that I hate and torture academics are fallen properties. Given GOOD_{O}, therefore, the relevant properties do not make a positive contribution to my overall intrinsic value. Indeed, the relevant properties make a negative contribution to my overall intrinsic value. So the refined theory can overcome even the present, seemingly formidable, variant of the "What if God Were Different?" objection.

But leave that to the side. Let us grant, for the sake of argument, that there is some way for God and me to torture academics without causing any decrease in the degree to which they are similar to Him and, hence, without causing a decrease in the degree to which they are fundamentally intrinsically good. This leads to the second reply: In order to obtain the desired result, BAD_I may be revised to include other sorts of badness besides decreases in fundamental intrinsic value. In *Finite and Infinite Goods*, Adams (1999, p. 93-101) offers an account of welfare. He analyzes welfare in terms of pleasure, pain, and (what I am calling) fundamental intrinsic goodness[11]. A natural revision of the theory, then, would incorporate Adams' discussion of welfare into the account of instrumental badness. This could be done by reformulating BAD_I so that it includes not just decreases in a subject's fundamental intrinsic goodness but also decreases in a subject's welfare. Then my being such that I hate and torture academics would contribute to the degree to which I am instrumentally bad. So condition (v_1) of FALLEN would be satisfied. So, by BAD_2, those properties would make a negative contribution to the degree to which I am overall intrinsically valuable[12].

1.7 The "What if Adams Believed he was God?" Objection

Suppose that Adams believed he was God[13]. If he were to believe this, then the

[11] Adams' discussion of welfare appears in Chapter Three of *Finite and Infinite Goods*. For similar theories see Feldman (2002) and (2004) as well as Kagan (2009b).

[12] Keep in mind here that fundamental intrinsic value and overall intrinsic value are distinct.

[13] Several people have presented me with one variant or another of the following objection: "God is infinite and everything else is finite. So nothing is more similar to Him than anything else. Consider the case of beliefs. Suppose that I have 5 true beliefs and you have a billion. In that case we are both equally far from God's omniscience. So, neither of us is more similar to God than the other with respect to being omniscient." It is important to see that this objection is mistaken. Consider similarity to God with respect to the avoidance of false beliefs. God has 0 false beliefs. Suppose that I have 40 billion false beliefs and that you have 13 false beliefs. 13 is much closer to 0 than 40 billion. So you are much more similar to God with respect to the avoidance of false beliefs than I am. Consider similarity to God with respect to the probability that one's beliefs are true. The probability that God's beliefs are true is always 1. The probability that our beliefs are true varies. But suppose that on this occasion the probability that my beliefs are true is .2 and yours is .75. A probability of .75 is much closer to 1 than a probability of .2. So you are much more similar to God with respect to the probability that one's beliefs are true than I am. Consider being similar to God with respect to, not the number of beliefs He has, but instead each particular belief that He has. Suppose I believe the proposition '2 + 2 = 4'. I am not more similar to God than my counterpart who does not believe this proposition with respect to having the same number of beliefs that He does. This is for the very reason that the objector points out. But I am more similar to God, than my

degree to which he is similar to God would be increased. So, given GOOD, the degree to which Adams is fundamentally intrinsically good would be increased. And, given GOOD_O and BAD_2, if Adams were to believe he is God, then the degree to which he is overall intrinsically valuable would be increased. But the degree to which Adams is overall intrinsically valuable would not be increased if he were to believe that he is God. So, GOOD_O is false[14].

This formulation of the objection is not persuasive. The premise 'If Adams were to believe he is God, then the degree to which he is similar to God would be increased' is false. To see this, consider David Lewis' (1973, 1979) analysis of counterfactuals[15]. Let @ be the actual world. Then contrast two possible worlds W1 and W2.

W1: Adams believes 'I am God' at W1. The behavior of Adams at W1 does not differ in any way from the behavior of Adams at @. At both worlds he is a philosopher, he is the author of *Finite and Infinite Goods*, he goes to church and receives the Eucharist, he has exactly the same conversations and friends, and so on. The beliefs of Adams at W1 are almost identical to the beliefs of Adams at @. The only difference

counterpart, in this sense: God and I share the property of having the belief '2 + 2 = 4'. Aside from beliefs: Consider the property of being a person. God has the property of being a person. You and I have the property of being a person. You and I sharing this property with God makes us more similar to Him than, say a rock or a book, is. Consider being actualized or existing. God exists and is actual. You and I exist and are actualized. So in this respect we are more similar to Him than some merely possible or non-existent object.

[14] Adams (1999, p. 32) very briefly discusses this objection.

[15] Lewis' analysis is this:
A counterfactual 'If it were the case that *A*, then it would be the case that *C*' is (non-vacuously) true at a possible world, *W*, if and only if there exists some (accessible) possible world at which *A* and *C* are true that is closer to *W* than any possible world at which *A* is true and *C* is false.
Closeness, for Lewis, is a similarity relation between possible worlds governed by the following system of weights and priorities:
(1) It is of the first importance to avoid big, widespread, diverse violations of law.
(2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
(3) It is of the third importance to avoid even small, localized, simple violations of law.
(4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

between the two is that Adams at W1 has the extra belief that 'I am God'. There is no belief that Adams at W1 has that Adams at @ does not have. For every belief that Adams at @ has, including 'Adams is not God', Adams at W1 also has it.

W2: Adams believes 'I am God' at W2. The behavior of Adams at W2 is very different from the behavior of Adams at @. Instead of going into philosophy, Adams at W2 joined a cult. That is how he came to believe he is God. His personality is very different. He has long hair and he frequently enjoys marijuana cigarettes. He does not have very many of the same friends that he has at @. He does not do any serious philosophy or go to church. The beliefs of Adams at W2 are very different from the beliefs of Adams at @. Adams at W2 has a lot of crazy and cultish false beliefs. He tries to fit together his belief that he is God with his other beliefs. Adams at W2 has many more false beliefs, such as 'Adams is God' and far fewer true beliefs, such as 'Adams is not God', than Adams at @.

Given Lewis' criteria for determining closeness between possible worlds, W2 is closer to @ than W1[16]. So the relevant counterfactual must be evaluated at a W2-like world rather than a W1-like world. But Adams is less similar to God at W2-like worlds

[16] To see this reason as follows: According to Lewis, avoidance of a big widespread and diverse violation of @'s laws has more weight in determining closeness to @ than matching @ with respect to particular matter of fact. Furthermore, Adams' belief that 'I am God' at W1 requires such a violation of the psychological laws governing beliefs with respect to @. Here is why: No one at @ has beliefs that are so dramatically causally isolated from each other and from behavior in the way that Adams' belief at W1 is. Adams' belief at W1 was not caused by anything. For each instant of time at which Adams believes 'I am God' at W1, he is held back from forming closely related beliefs such as 'Adams is God' and from dropping beliefs that are obviously inconsistent with 'I am God' such as 'Adams is not God'. Adams' belief never shows up in his behavior. Not even his psychotherapist can tell that Adams believes 'I am God'. Not even Adams believes 'Adams believes he is God.' At W2, on the other hand, Adams' belief conforms to the psychological laws of @. His belief is caused by a thorough brainwashing combined with a series of ecstatic marijuana induced rituals at the cult compound. His belief causally interacts with his other beliefs producing and being reinforced by them. His belief has a causal influence on his behavior. Though the result is much less of a match with @ with respect to particular matter of fact, it avoids a violation of the psychological laws of @. So W2 is closer to @ than W1.

than he is at @[17]. So the relevant counterfactual is false. So this formulation of the objection is unpersuasive.

1.8 A Different Formulation of the "What if Adams Believed he was God?" Objection

There is another, more plausible, way of formulating the objection. The real worry is not about whether the counterfactual in question is true or false. The worry is instead that there is some W1-like possible world, no matter how remote from @ with respect to psychological laws, at which Adams believes 'I am God' and everything else is left the same as it is at @. At W1 the degree to which Adams is similar to God is greater than it is at @. So if GOOD_O is true, then the degree to which Adams is overall intrinsically valuable at W1 is greater than the degree to which Adams is overall intrinsically valuable at @. But the degree to which Adams is overall intrinsically valuable at W1 is not greater than the degree to which Adams is overall intrinsically valuable at @. So GOOD_O is false.

While this formulation of the objection is more persuasive than the first, it is unsound. By believing 'I am God', Adams at W1 gains a belief in common with God. But its truth-value is relative to the person who believes it. The belief is false with respect to Adams and it is true with respect to God. So, while Adams at W1 shares a belief with God that Adams at @ does not, Adams at @ is more similar to God than Adams at W1 with respect to the avoidance of false beliefs. This property that Adams at @ keeps and Adams at W1 loses has more weight in determining similarity to God than

[17] To see this, reason as follows: Adams at W2 shares one new belief with God, namely 'I am God', that Adams at @ does not have. But Adams at W2 gains many crazy and false beliefs that God does not have such as 'Adams is God' and 'God should leave the compound sometime today and head over to the grocery store to get some stuff for the fridge' and he loses many sane and true beliefs that he has in common with God at @ such as 'Adams is not God' and 'God was not born in the 20th Century'. So at W2 Adams is less similar to God than he is at @.

the belief that Adams at W1 gains in common with God does. So Adams at W1 is not more similar to God than Adams at @.

The controversial premise in the argument above is this: 'Increasing match to God with respect to the avoidance of false beliefs has more weight than believing 'I am God' in determining similarity to God'. To support this premise, consider something David Lewis (1986, p. 43) says about the role that similarity plays in his analysis of counterfactuals:

> The thing to do is not to start by deciding, once and for all, what we think about similarity of worlds, so that we can afterwards use these decisions to test [my analysis]. What that would test would be the combination of [my analysis] with a foolish denial of the shiftiness of similarity. Rather, we must use what we know about the truth and falsity of counterfactuals to see if we can find some sort of similarity relation—not necessarily the first one that springs to mind—that combines with [my analysis] to yield the proper truth conditions. It is this combination that can be tested against our knowledge of counterfactuals, not [my analysis] by itself. In looking for a combination that will stand up to the test, we must use what we know about counterfactuals to find out about the appropriate similarity relation—not the other way around.

Given how influential Lewis' analysis is, I think that I can help myself to the same strategy when testing Adams' theory of intrinsic value. Let us use what we know about intrinsic value to discover the appropriate similarity relation to God. What we find, if we do that, is the similarity relation that assigns great importance to increasing one's match to God with respect to the avoidance of false beliefs and assigns little or no importance to increasing one's match to Him with respect to His *de se* beliefs such as 'I am God'. Combined with Adams' theory this similarity relation delivers the correct judgment: Adams at W1, that is Adams at a world at which he believes 'I am God' but everything else, other than the psychological laws, is left the same as it is at @, is not more similar to God than Adams at @. So the "What if Adams believed he was God?" objection is unsound.

1.9 Conclusion

When modified in appropriate ways, Adams' theory is able to deal with a number of natural, seemingly formidable, objections. Adams' *Finite and Infinite Goods*, I conclude, is suggestive of an interesting theory of intrinsic value that is worthy of critical attention.

CHAPTER 2

NATURALISM AND MORAL REALISM

Robert Adams, Alvin Plantinga, and Michael Rea have defended some objections to Naturalist Moral Realism (NMR). In this paper I describe a form of NMR and argue that it evades each of their objections. Some of the objections are simply unpersuasive. Others can be reformulated into interesting objections. These more interesting reformulations, however, are all dependent on objections to NMR already widely discussed in the literature. A consequence of this dependence is that Adams, Plantinga, and Rea's objections contribute nothing to the debate about whether naturalism can accommodate moral realism.

2.1 The Causal Theory of Reference

Proponents of NMR often appeal to the Causal Theory of Reference (CTR). CTR includes an account of *reference grounding*. In the case of names, CTR holds that:

A name, 'N', is grounded in an object, N, only if a subject points to N and utters some sentence like 'That is N'.

In the case of natural kind terms, CTR holds that:

A natural kind term, 'K', is grounded in a natural kind, K , only if a subject points to a sample of K and utters some sentence like 'That is K'.

Of course, these necessary conditions are not, by themselves, sufficient. More needs to be said about further conditions that must be met by the speaker and by the speaker's environment to ensure the successful grounding of '*N*' in *N* and of '*K*' in *K*. While more has been said[18], and progress has been made, it is still very difficult to specify exactly what else must be added.

In addition to an account of reference grounding, CTR also includes accounts of *reference borrowing* and *reference change*. Concerning the account of reference borrowing, proponents of CTR hold that a subject need not be present at a successful grounding of a term in an object or property in order for that subject to use that term to refer to that object or property. All that is required is that, first, the subject has witnessed someone else successfully use that term to refer to the relevant object or natural kind and, second, the subject uses that term with the intention of referring to whatever that other person referred to when he or she used that term.

Concerning the account of reference change, consider Evans' famous 'Madagascar' example. At one time 'Madagascar' referred to mainland Africa. Now it refers to an island off the coast of mainland Africa. If one act of reference grounding determines the referent of a term once and for all, then 'Madagascar' would still refer to mainland Africa. But 'Madagascar refers to an island and not to mainland Africa. So an account of reference change is needed. Of course, we are here talking about natural kind terms in addition to names. But Evan's lesson about Madagascar applies to natural kind

[18] What more has been said? This will become important later. But for now here, in slogan form, is a sample of some of the further conditions have been added to the simple account of grounding: Devitt and Sterelny (1999) have added the condition that the speaker who grounds the term associates that term with a description that determines one and only one of the natural kinds to which the relevant sample object belongs. Stanford and Kitcher (2000), have added among others the condition that the speaker who grounds that term associates that term with a range of samples, a range of foils, and a conjunctive sentence that is such that each sample satisfies each of the conjuncts and each foil fails to satisfy at least one of the conjuncts. Boyd (2003) holds that a term refers to a natural kind with respect to a disciplinary matrix only if that natural kind, together with the use of that term, explains the achievements of the practitioners within that disciplinary matrix. Soames (2002) adds the condition that the speaker has various dispositions that together with the speaker's act of dubbing help fix reference. Brown (1998) and Jylkka (2008) have added the condition that a speaker that grounds that term has a recognitional capacity, or a recognitional component, that will eventually be discovered by psychologists.

terms and moral terms as well. The reference of a name or natural kind term changes when a weighted most of that term's groundings are in a new object or natural kind[19].

2.1.1 Extending CTR from Scientific Terms to Moral Terms

Many proponents of NMR propose extending CTR from names and natural kind terms to moral terms. Two motivations for this extension are often cited. First, it is sometimes alleged that CTR allows the naturalist to deal with objections to scientific realism[20]. Some naturalists interpret Putnam and Kuhn as having argued that, given a descriptivist theory of reference for natural kind terms, naturalism cannot accommodate scientific realism. CTR, on the other hand, is alleged to overcome these Putnamian and Kuhnian objections. Since the authors Adams, Plantinga and Rea target see scientific inquiry as the only legitimate form of inquiry, and since they hope to include moral realism under the umbrella of scientific realism, they are motivated to extend CTR to moral terms.

Second, it is sometimes alleged that extending CTR to moral terms allows the naturalist to escape Moore's Open Question Argument. Proponents of NMR sometimes

[19] The account of reference change, though widely adopted, is not present in Kripke and Putnam's initial formulation of CTR. It is due to Devitt. See, for example, Devitt and Sterleny () and Reimer ().

[20] Kornblith (2006, 16) makes the point this way:

Putnam's account of the reference of natural kind terms was part of an attempt to offer a realist philosophy of science…. The traditional descriptions theory of reference made it impossible to understand the possibility of disagreement between individuals committed to different theories…. Putnam's argument against a descriptions-based theory of reference for natural kind terms was designed to show that such a theory fails to explain a certain real phenomenon, namely, progress in science….

Laurence and Margolis (2003, 271-2) also take this line:

Ironically, Putnam's main contribution to the development of causal theories isn't Twin Earth but rather his observations concerning purported a priori analyses and how they don't mesh with… the history of science…. One of the advantages of causal/anti-descriptivist theories is that they do.

The same thought is expressed by Kuhn (2000):

To avoid [objections to scientific realism] and related problems from other sources, many philosophers have in recent years emphasized… the so-called causal theory of reference developed by Kripke and Putnam. It is rooted firmly in possible world semantics, and its expositors resort repeatedly to examples drawn from scientific development.

interpret Moore as showing that, given a descriptivist theory of reference, moral properties cannot be identical to natural properties. They go on to claim that, unlike descriptivist theories of reference, CTR allows moral properties to be identical to natural properties[21]. Since moral realism is sometimes formulated in terms of properties and since natural properties are the only properties acceptable within naturalism's austere ontology, they are motivated to extend CTR to moral terms.

Proposing such an extension might seem puzzling. There are important respects in which moral terms are disanalogous with names and disanalagous with natural kind terms. First, names are referring terms. Many moral terms, however, are used primarily as adjectives. Consider sentences such as 'That man is good' and 'The oil spill is not good'. In these sentences 'good' is used as an adjective. In such cases 'good' does not refer to anything. Since moral terms like 'good' are not primarily referring terms, it is unclear how CTR is supposed to be extended to such terms. So when authors say that CTR extends to moral terms, they must not be suggesting that it is extended to moral terms that are used as adjectives[22].

Second, many natural kind terms refer to concrete stuff. And, in some cases, a natural kind term refers to the property of being certain concrete stuff. So, for example, 'water' refers to H_2O and 'waterocity' refers to the property of being H_2O. Consider, however, referring moral terms such as 'goodness'. 'Goodness' refers to a property and

[21] Although, as Feldman (2005) points out, people who talk this way often simply misunderstand Moore's Open Question Argument. Feldman (2005), for example, says this:
In recent discussions of OQA [the open question argument], commentators sometimes suggest that discoveries concerning rigid designators, natural kind terms, the causal theory of reference, etc., have shown us that OQA is little more than a quaint relic. Saul Kripke and Hilary Putnam are often mentioned. Putnam himself seems to endorse this line of thinking…. Putnam starts out by saying that Kripke's ideas have a devastating impact on Moore's argument. He then mentions in passing that the most Moore's argument establishes is that 'good' is not synonymous with 'we desire to desire'. But what is astonishing here is that *this is precisely the conclusion that Moore was concerned to establish!*

[22] Soames (2000) makes a similar point about natural kind terms.

not to stuff. And there is no concrete stuff that can be identified with all and only things that instantiate goodness. So ‘goodness’ cannot refer to the property of being certain concrete stuff. So when authors say that moral terms are analogous to natural kind terms, they must not mean that moral terms and moral properties are analogous to natural kind terms and natural kinds in the respects just specified.

In light of these disanalogies, perhaps authors who suggest extending CTR to moral terms have the following extension in mind: First, CTR applies only to moral terms that are referring terms such as ‘goodness’ and not to moral terms that are adjectives such as ‘good’. Second, moral terms are analogous to natural kind terms, not in the sense that they refer to concrete stuff or the property of being certain concrete stuff, but in the sense that the referents of moral terms are fixed by ostension. Extended in this way, CTR holds that:

> A (referring) moral term, ‘M’, is grounded in a moral property, M, only if a subject points to a sample of, M, and utters some sentence like ‘That is M’.

Even more so than in the case of names and natural kind terms, more needs to be said about what, in addition to pointing and uttering, is sufficient for grounding ‘M’ in M. As we will see, this is a difficult problem to address. But perhaps authors who promote extending CTR to moral terms have the following sketch in mind: Suppose ancient moral theorists find that Nature has disposed them to classify a number of very different items together. They have a list of samples that belong in this class. They see two people in deep conversation, that the harvest was plentiful this year, and the pattern of the stars in the sky. These, along with many other items, are samples. The ancient moral theorists have a list of foils that do not belong to this class. They see a young man being eaten by a wild beast and that a pile of refuse is now an object of worship to many. These, along with many other items, are foils. Now, suppose that one of the moral theorists points to one of the samples with the intention of identifying whatever it is that the samples have in common and the foils lack. Suppose the moral theorist then utters some sentence like 'That is goodness'. Somehow the ancient moral theorist's mental states and the ancient moral theorist's environment combine with the act of pointing and uttering to successfully ground 'goodness' in a unique property.

2.1.2 Reference, Truth, and Identity

Proponents of NMR sometimes supplement CTR with a claim about truth and a claim about identity. Consider the natural kind term 'water'. CTR, combined with various causal and historical assumptions, is said to deliver the judgment that 'water' refers to H_2O. Establishing that the referent of 'water' is H_2O is supposed to be because proponents of NMR often supplement CTR with the following claims:

TRUTH: The truth-value of a sentence depends on the referents of the reference bearing terms in that sentence and on the structure of that sentence.[23]

IDENTITY: If a term, 's1', refers to some stuff, s2, then the property of being s1 is identical to the property of being s2.[24] [25]

[23] Kuhn (2000) attributes TRUTH to Kripke and Putnam. He says this:
[M]any philosophers have in recent years emphasized that truth values depend only on reference, and that an adequate theory of reference need not call upon the way in which the referents of individual terms are in fact picked out.
Devitt and Sterelny (1999, p. 21-2) endorse TRUTH as well. They say this:
What is it about a word that affects the truth conditions of sentences containing it…? The obvious answer is: its *referent*…. [T]he truth condition of a sentence depends on its syntactic structure and the referents of its words…. Given reference and structure, truth conditions are determined.
[24] Scott Soames (2008, p. 16) embraces IDENTITY. He says this:
[*B*]*eing water* must be the property *being H2O*, which, in turn, must be the meaning of the predicate 'is water' (but not of the predicate 'is H2O').
LaPorte (1996, pp. 1,) who himself rejects IDENTITY notes that it is "widely popular". He says this:
I am inclined to say that this identity statement and others like it are false. It is an empirical fact, perhaps, that all water is H20 and all H20 is water. But the claim that the one is identical to the other needs more than empirical support. Essentialists have looked for the needed support in thought experiments alleged to illustrate that speakers would call all and only H20 'water' in any possible world, provided only that 'water' and 'H20' are in fact coextensive.
Adams (197?) and (1999) endorses IDENTITY in the following passages:

I take it that TRUTH is supposed to work as it does in the following example: Consider the sentence 'There is water in my cup'. Whether this sentence is true depends on whether the stuff in my cup is H_2O. If that stuff is H_2O, then the sentence in question is true. If not, then the relevant sentence is false.

To be made plausible, however, TRUTH will require elaboration, clarification, and refinement. In particular, the dependence relation discussed in TRUTH will need to be spelled out in detail. The relevant sort of dependence cannot, for instance, yield an interpretation of TRUTH that implies this:

DEP: For any reference bearing term in a sentence, if the reference of the term is changed, then the truth-value of that sentence is changed.

DEP is clearly false. Consider a sentence[26] like 'Water is a liquid and 2+2 = 5.' This sentence is false no matter what 'water' refers to. So if the reference of 'water' is changed, then contrary to DEP, the truth-value of the relevant sentence stays the same. Or, consider the sentence 'Water is a liquid'. This sentence is true. Suppose the reference of 'water' is not to H_2O but to some other liquid stuff. Then the relevant sentence is still true. Reference has changed, but truth-value has not changed. So DEP is false. TRUTH will, therefore, require further refinement. And, in fact, it requires more refinement than I can offer here. Ultimately, my goal is neither to defend nor to attack NMR. I only hope to advance discussion of this topic by clarifying NMR and by

[25] Amie Thomasson seems to endorse TRUTH and IDENTITY like claims. She says this:
[I]t is a platitude about reference that (for most general terms *K*, assuming the term exists), *K* refers if and only if Ks exist. This link between the fundamental rules of use for 'refers' and 'exists' enables us to move up and down the semantic slide from talking (in the metalanguage) about whether a term refers to speaking (in the object-language) about whether or not the relevant things exist…. [We can] shift between using terms in talking about whether or not entities of a given sort exist and mentioning those terms in discussing whether they refer. This shift is crucial to enabling us to assess the truth-value of existence claims.

[26] Thanks to Fred Feldman for suggesting this example.

demonstrating that it can evade the objections raised by Adams, Plantinga, and Rea. Furthermore, none of the objections offered by Adams or Plantinga or Rea target TRUTH. So I will leave it to proponents of NMR to hammer out the precise details of TRUTH. For the purpose of this paper, assume that they have.

IDENTITY, although very popular, is certainly not uncontroversial. I will discuss this more in the section on Rea's objection. For now, I will just explain how IDENTITY is supposed to work and why, together with CTR, IDENTITY delivers interesting results. Given IDENTITY and given CTR's judgment about 'water' and H_2O, the property of being water and the property of being H_2O are identical properties. English speakers may associate a different description with 'water' than they do with 'H_2O'. People may think that being water and being H_2O are distinct properties. But, it is claimed, they are not. The property of being water and the property of being H_2O are one and the same property.

As we have seen, when supplemented with TRUTH and IDENTITY, CTR generates interesting results. Suppose, in addition, that the sketch given above is correct. CTR, combined with causal and historical assumptions, delivers the judgment that a moral term like 'goodness' refers to some property P. Then the following two claims are alleged to be true:

TRUTH: The truth-values of sentences that include 'goodness' depend on P and on the structure of those sentences.

IDENTITY: Goodness is identical to P.

Consider, then, a sample event or object or any other sort of stuff that we take to instantiate *goodness*. Now consider sentences such as 'The oil spill does not instantiate goodness' or 'That person instantiates goodness'. These sentences depend on whether the relevant sample instantiates P and on the structure of those sentences. Furthermore, goodness and P are identical properties. People may associate a different description with ‘goodness’ than they do with ‘P’. They may think that goodness and P are distinct. But these properties are not distinct.

It will be important to evaluate the truth-values of sentences containing moral terms that are not referring terms. Suppose we want to find the truth-value of a sentence containing non-referring moral terms such as ‘That is a good man.’ Such sentences will have a paraphrase in which the non-referring moral term is replaced with a referring moral term. For example, ‘That is a good man’ has as a paraphrase ‘That man instantiates goodness’. The truth-value of ‘That is a good man’ is identical to the truth value of its paraphrase.

2.2 Moral Realism

There are different formulations of moral realism in the metaethical literature. Some formulations are in terms of which sentences are true and which sentences are false. Others are in terms of which properties exist or are instantiated. Others are a combination. I will discuss each and then settle on a combined account of moral realism.

The Sentence Formulation: Many authors understand Ethical Nihilism as a claim about which moral sentences are true and which moral sentences are false[27]. If reference, together with structure, determines the truth-values of sentences, as TRUTH indicates,

[27] Maitzen (1998, p. 357) says that Ethical Nihilism is the claim “that no ethical sentence, standardly construed, is true. Pigden (2007, p. 451) says that “meta-ethical nihilism needs to be reformulated. I suggest the following: All nonnegative atomic moral judgments are false. This requires elucidation… an atomic moral judgment [is] a proposition….” Schaffer-Landau (2004, pp. 8-) sometimes formulates Ethical Nihilism in this way. He says “Moral nihilism denies that there are any moral truths…. This is an error theory... According to such a view, moral terms have descriptive meaning and we have moral beliefs, but they are uniformly false.”

then determining the referents of moral terms will allow us to determine which moral sentences are true and which are false. If at least some moral sentences (of a certain sort) are true, then Ethical Nihilism is false. If no moral sentences (of a certain sort) are true, then Ethical Nihilism is true[28]. Perhaps we will not go too far astray if we suppose that those who understand Ethical Nihilism in terms of the truth-values of sentences will also understand other metaethical positions, such as moral realism, in terms of the truth-values of sentences. Indeed, some authors explicitly understand moral realism in this way. According to Sayre McCord[29], for example, "Moral realism is not a particular substantive moral view nor does it carry a distinctive metaphysical commitment over and above the commitment that comes with thinking moral claims can be true or false and some are true." Other authors may want to add the condition that the truth-values of moral sentences are objective. Sayre-McCord, himself, seems to endorse this extra condition elsewhere. Combining these two conditions we are in a position to consider what we may call "the Sentence formulation of MR" or "SMR":

SMR: Some moral sentences (of the relevant sort) are true and the truth-values of those sentences are objective.

Suppose that authors such as Sayre-McCord are correct in formulating Moral Realism as SMR. Then determining the referents of moral terms will allow us to determine the truth-values of moral sentences and whether those truth-values are objective. If it turns out that some moral sentences are true and that the relevant truth-values are objective, then moral realism is true. If not, then moral realism is false.

[28] I say "of a certain sort" because one might think that both 'Torturing babies is wrong' and 'It is not the case that torturing babies is wrong' are moral sentences. One or the other sentences has got to be true. So, Ethical Nihilism, on this reading, would be too easily refuted. Some authors (e.g. Maitzen (1998)) deal with this problem by saying that one or the other sentences is not really moral. Other authors (e.g. Pigden (2007)) deal with this problem by saying that both sentences are moral but Ethical Nihilism is true if every moral sentence of a certain sort is false.

[29] See Sayre McCord (2009).

The Property Formulation: Other authors understand MR as the claim that there are objective moral properties[30]. Call this "the Property formulation of Moral Realism" or "PMR":

PMR: There are moral properties and they are objective.

Suppose that Moral Realism is best formulated as PMR. Then CTR, together with causal and historical assumptions about speakers and their environments, will determine whether and to what moral terms refer. Consider, for example, the moral term 'wrong'. Suppose that CTR delivers the judgment that 'wrong' fails to refer. Then, by IDENTITY, wrongness doesn't exist. So wrongness is not an objective moral property. So Moral Realism, with respect to wrongness, is false. Suppose instead that CTR delivers the judgment that 'wrong' successfully refers to some property but that the relevant property is not objective. Then, by IDENTITY, wrongness is identical to some property that is not objective. Assume the same is true of other moral terms and the properties to which they refer, then there are moral properties but they are not objective. Then PMR is not true. If, however, 'wrong' successfully refers to some property and that property is objective, then PMR is true with respect to wrongness and, given that other moral terms successfully refer to objective properties, PMR is true in general.

A Combined Formulation: Some authors offer an account of Moral Realism that combines SMR and PMR. Plantinga, for example, offers such an account. I think this is the right line to take. It seems to me that the spirit of Moral Realism is not fully captured by either SMR or PMR. Both seem necessary for the truth of Moral Realism. But neither seems sufficient. In addition, a more full account of Moral Realism should add a dependence relation between SMR and PMR along with the requirement that the moral

[30] Rea Plantinga Boyd Adams

judgments of ordinary speakers cannot deviate too far from which moral sentences are actually true and which are false. I will take ‘Moral Realism’ to apply only to theories that satisfies all of these conditions:

Moral Realism: PMR is true, SMR is true, the truth of SMR depends on the truth of PMR, and a large body of ordinary judgments about the truth-values of moral sentences match the actual truth-values of moral sentences.

I have already discussed naturalist strategies for accommodating PMR and SMR. If CTR delivers the judgment that moral terms refer to objective natural properties, then the proponent of NMR can satisfy the additional conditions that I have said must be satisfied by any version of Moral Realism. Since the truth-values of moral sentences depend on the structure of those sentences and the properties referred to by moral terms and since, given PMR, those properties are moral terms, the truth of SMR depends on the truth of NMR. Suppose in addition that the truth-values of moral sentences match the truth-values that we ordinarily assign to moral sentences. Then the proponent of NMR can accommodate Moral Realism.

2.3 Adams’ First Objection

Robert Adams’ important book *Finite and Infinite Goods* contains some interesting critical discussions of NMR. One chapter is devoted entirely to NMR. NMR is discussed critically in other substantial chapters as well. Here I want to consider two passages in which Adams raises some problems for NMR. One alleged problem concerns the proponent of NMR’s use of CTR. Adams argues that CTR cannot accommodate the standard line that ‘water’ determinately refers to H_2O rather than other stuff. Adams further argues that any modification to CTR that would enable it to deliver the relevant judgment would yield widespread non-objectivity about “claims that link metaphysics and semantics on any topic" including ethics. But Moral Realism requires that moral properties be objective. So if this objection is sound, it would be a serious problem for the proponent of NMR. Here is the passage in which Adams discusses this objection:

> Suppose we find on another planet a substance that shares the observable common properties of water (colorless, odorless, tasteless, it quenches thirst, forms solid, liquid, and gaseous states at the same temperatures as water, etc.), but is not H_2O; instead it is constituted in an analogous way of two elements unknown on Earth, so that we may call it, for present purposes, $\Phi_2\Psi$. Is $\Phi_2\Psi$ water? The received answer is that it is (necessarily) not, and I am prepared to agree; but what makes it the right answer? Our seventeenth-century ancestors, who knew nothing of the periodic table, let alone H_2O or $\Phi_2\Psi$, used the word 'water', very competently, in the same sense as we do. What is determined about the nature of water by the sense of 'water', as we share it with them, can hardly be more than that it accounts causally for the observable properties of water. In my example H_2O and $\Phi_2\Psi$ may be assumed to have analogous structures which are causally relevant in analogous ways to their shared observable properties. Why not say, then, that it is not being H_2O, but a structural property common to H_2O or $\Phi_2\Psi$, that accounts causally for the observed properties of water, even on planet Earth[31]?

What exactly is the objection Adams means to be suggesting here? As I see it, there are two possibilities. First, Adams might be arguing that CTR is unable to judge that the actual English term 'water' refers to H_2O rather than the structure common to H_2O and $\Phi_2\Psi$. If this is his objection, then perhaps his reasoning can be reconstructed as follows: A sample of H_2O is also a sample of the structure common to H_2O and $\Phi_2\Psi$. According to CTR, an act of reference grounding is successful if and only if a subject points to a sample of some natural kind, K, and utter some sentence like 'That is K.' But there is nothing about the subject's pointing and uttering that determines that 'water' refers to H_2O rather than the structure common to H_2O and $\Phi_2\Psi$. So if CTR were true, then 'water' would not refer to H_2O rather than the structure common to H_2O and $\Phi_2\Psi$.

This is a very interesting objection. But this is a specific instance of an objection already widely discussed in the literature on CTR—the Qua Problem. One need not point

[31] 23-4

to a fanciful thought experiment about $\Phi_2\Psi$ or any "shared common structure causally relevant in an analogous way to the observed properties of water" to generate the problem. One need only note that a sample of H_2O is also a sample of many other things. It is a particular sample of H_2O. It is, at the time of pointing, a sample of a particular temporal slice of that particular sample of H_2O. It is a liquid. It is an instance of the disjunctive property being either H_2O or an electronic musician. Nothing about an act of pointing and uttering, by itself, determines that the subject is pointing to the sample *qua* sample of H_2O rather than the sample *qua* sample of any of the other things it is a sample of. So CTR will require some tinkering if it is to deliver the judgment that 'water' refers to H_2O rather than any of the other things the relevant sample is a sample of.

Over the years, proponents of CTR have introduced modifications and additions to deal with this problem and the proponents of the Qua Problem (often the proponents and critics are the same authors) have criticisms of those modifications[32]. Here I am only

[32] The dialectic has gone this way: Proponents of CTR attempt to avoid the Qua Problem by positing features of the mental states of speakers that single out one of the natural kinds instantiated by the relevant sample when those speakers perform an act of grounding. The problem with this move is that proponents of CTR end up attributing to such speakers mental states that no human has. These variants of CTR then become subject to the same sorts of objections that CTR's traditional rival-the Descriptivist Theory-suffers from—namely, the so called "Problem of Ignorance and Error." The resulting theories, then, suffer from both the Qua Problem and the Problem of Ignorance and Error. This point is made in many places but perhaps most notably in Devitt and Sterelny (1999) and Jylkka (2008).

One might think that prioritizing naturalness in reference grounding and appealing to the naturalness of natural kinds would help the proponent of NMR. In addition to the mental states of speakers, it is the "naturalness" of the properties instantiated by a sample that help to determine reference. And natural kinds are more natural than other candidate referents.

This is an interesting line to take in the case of natural kind terms. The appeal to naturalness certainly helps explain why CTR delivers the judgment that 'water' is grounded in H_2O rather than in being H_2O or an electronic musician. But I don't think this move will help the proponent of NMR. The problem is that the properties proponents of NMR allege to be identical to moral properties are very different than the standard examples of natural kinds offered by naturalists.

Consider, for example, Richard Boyd's homeostatic property cluster theory. As Rubin (2008) points out, the property cluster that Boyd identifies with goodness does not have the kind of metaphysical structure that is characteristic of all of the other homeostatic property clusters that Boyd takes to be natural kinds. So if one adopts Boyd's theory of natural kinds and one privileges naturalness in one's account of reference grounding, then, since the property Boyd identifies with goodness is, by his own standards, not a natural kind, such privileging does not alleviate the Qua Problem.

making two points: First, if this is the objection Adams means to be suggesting, then his objection is unoriginal. It was widely discussed in the literature on CTR decades before the publication of *Finite and Infinite Goods*. Moreover, Adams does not cite or engage with any of those proposed modifications of CTR. Second, whatever else Adams' objection may be, it cannot turn out that its success depends in some implicit way on the success of the Qua Problem. If it does, then the real debate is about whether the Qua Problem is successful and not about Adams' own objection. And that debate had been raging long before *Finite and Infinite Goods* first appeared.

So what is the other objection Adams might be suggesting in the relevant passage? Perhaps the idea is that CTR judges that before Adams' planet is discovered the referent of 'water' is H_2O and after the planet is discovered the referent of 'water' is the structure common to H_2O and $\Phi_2\Psi$. But the discovery of the new planet, the objection might proceed, should not carry with it a change in the reference of 'water'. If this is what Adams has in mind, then I can see two ways that this sort of objection might go.

Here is the first way: Suppose that an English speaker lands on Adams' planet. She points to the watery stuff on that planet and says something like 'Look! They have

In addition, some proponents of NMR identify goodness with a complex disjunctive property. Brink (1989, p. 156-60), for example, does this. However, disjunctive properties, even very simple ones, are paradigm examples of properties that fail to count as natural kinds. So the move to naturalness is not going to help.

Other proponents of NMR, such as Railton (1986) identify goodness with bizarre counterfactual properties such as being what would satisfy what our desires would be if we had "unqualified cognitive and imaginative powers, and full factual and nomological information about [our] physical and psychological constitution, capacities, circumstances, history, and so on." These farfetched counterfactual properties are very different than the sorts of properties that naturalists usually take to be natural kinds such as being H_2O, being a tiger, and being an electron.

So, given the properties that advocates of NMR actually identify with moral properties, prioritizing naturalness in CTR's account of reference grounding is going to hurt the prospects for a plausible account of NMR rather than help. Or, at the very least, the proponent of NMR owes us a plausible theory of natural kinds that includes the sorts of properties that they take to be identical to moral properties and that gives us an error theory that explains why the properties in question seem so different than the standard examples of natural kinds and yet still turn out to be natural kinds in spite of those differences.

water here too. That is water.' She will have grounded the term 'water' in a sample of $\Phi_2\Psi$. So there will have been groundings of 'water' in both samples of H_2O and a sample of $\Phi_2\Psi$. 'Water', therefore, must refer to something both sorts of samples have in common. What they have in common is not that they are all samples of H_2O. Instead, it is the structure common to H_2O and $\Phi_2\Psi$ that they have in common. So that common structure, rather than H_2O, is what 'water' refers to.

It is certainly counterintuitive to say that 'water', on the present interpretation of Adams' objection, refers to the structure common to H_2O and $\Phi_2\Psi$. Surely the proponent of CTR will want to say that the English word 'water' in such a case continues to refer to H_2O. Indeed, it seems clear to me that the proponent of CTR has the resources to maintain just this claim. According to the received formulation of CTR, a natural kind term refers to whatever natural kind is such that a weighted most of the groundings of that term are in that natural kind. So just one act of grounding 'water' in a sample of $\Phi_2\Psi$, or the structure common to H_2O and $\Phi_2\Psi$, would not change its reference. But if the vast majority of the groundings of 'water' came to be in samples of $\Phi_2\Psi$, or the structure common to H_2O and $\Phi_2\Psi$, then, the reference of 'water' would change from H_2O to $\Phi_2\Psi$, or the structure common to H_2O and $\Phi_2\Psi$. So, 'water' will continue to refer to whatever it has referred to in the past. A few groundings in a new natural kind will not change the reference of 'water.' One might wonder in the first place why groundings of 'water' in samples of H_2O secure reference to H_2O rather than the structure common to H_2O and $\Phi_2\Psi$. But as we have already seen, if this is the worry, it is just a special instance of the much discussed Qua Problem.

Here is the second way that this objection might go: Perhaps Adams is suggesting that once the planet has been explored thoroughly and people have traveled back and forth many times and there have been numerous groundings of ‘water’ in samples of $\Phi_2\Psi$, that CTR would deliver the judgment that ‘water’ refers to the structure common to H_2O and $\Phi_2\Psi$ rather than to H_2O. If this is his complaint, then I doubt that the proponent of CTR would be worried about this result. Note that proponents of CTR typically hold that there is a graded distinction between superficial properties of a sample and deeper properties of a sample. The watery-ness of a sample of water is more superficial than the structure common to H_2O and $\Phi_2\Psi$ which is itself more superficial than being H_2O. In cases where the same superficial properties are shared by samples with different deeper properties and where speakers frequently ground the same term in samples that share the superficial properties but vary with respect to the deeper properties, proponents of CTR want the theory to deliver the judgment that the relevant term refers to the more shallow property. Putnam (p. 241) says it this way:

> Some diseases, for example, have turned out to have no hidden structure (the only thing the paradigm cases have in common is a cluster of symptoms)…. An interesting case is the case of *jade*. Although the Chinese do not recognize a difference, the term ‘jade’ applies to two minerals: jadeite and nephrite. Chemically, there is a marked difference. Jadeite is a combination of sodium and aluminum. Nephrite is made of calcium, magnesium, and iron. These two quite different microstructures produce the same unique textual qualities! Coming back to the Twin Earth example, for a moment; if H_2O and XYZ had both been plentiful on Earth, then we would have been correct to say that there were *two kinds of ‘water’*. And instead of saying that ‘the stuff on Twin Earth turned out not to really be water’, we would have to say ‘it turned out to be the *XYZ kind of water*’. To sum up… the local water, or whatever, may have two or more hidden structures—or so many that ‘hidden structure’ becomes irrelevant and the superficial characteristics become the decisive ones.

My ability to summon within myself definitive NMRish intuitions about these cases is starting to get shaky. But if the case Adams means to be suggesting is one in which there are numerous groundings of ‘water’ in both samples of H_2O and $\Phi_2\Psi$ then I think the proponent of CTR would be happy to accept that water does not refer to H_2O, that maybe there are two types of water and ‘water’ refers to both. Or perhaps instead the proponent of CTR would want to hold that the structure shared by H_2O and $\Phi_2\Psi$ is the referent of ‘water’. So I don’t think the proponent of CTR, and hence the proponent of NMR under consideration, would be too troubled with this result

Another concern I have about Adams’ objection to NMR is that I’m not sure what Adams means by “the structure common to H_2O and $\Phi_2\Psi$.” One might think he means to be talking about the property of being composed of two atoms of one kind and one atom of another kind. But that can’t be it since there are other actual molecules with that structure that do not have the watery properties shared by samples of H_2O and samples of $\Phi_2\Psi$. Again, one might wonder why an act of grounding in a sample of H_2O is an act of grounding in H_2O rather than an act of grounding in molecules that are composed of two atoms of one kind and one atom of another kind since a sample of H_2O is a sample of both. But if that is the concern, then it is just the Qua Problem. And if it isn’t the Qua Problem that Adams is pressing, then the proponent of NMR, I think, has an easy story to tell about why ‘water’ refers to H_2O rather than to the structure common to H_2O and $\Phi_2\Psi$. The reason is that plenty of other samples of molecules are composed of two atoms of one kind and one atom of another kind, just like H_2O and $\Phi_2\Psi$, but they don’t have the relevant “superficial” watery properties had by samples of H_2O and samples of $\Phi_2\Psi$. So the Qua Problem notwithstanding, the extension of ‘water’ certainly doesn’t include those other samples of molecules. So it doesn’t refer to that common structure. That is why ‘water’ refers to H_2O rather than the structure common to H_2O and $\Phi_2\Psi$. Now, I am no chemist. Maybe there is something more to the structure of molecules than just how many atoms they have and what the comparative amounts of each of those atoms are. But there is certainly nothing in the passage in which Adams discusses this objection that points us to what that deeper structure might be. So without some story about molecular structure that will yield a structure common to H_2O and $\Phi_2\Psi$ that is not also had by some non-watery sample of molecules, I doubt that the proponent of NMR has anything to worry about.

2.4 Adams' Second Objection

Adams presents another objection to NMR. He describes a case in which two subjects disagree. But the proponent of NMR, Adams claims, cannot give a "realist" account of that disagreement. So NMR must be false. Here is what Adams says:

> Suppose Boyd's hopes for the progress of science to have been vindicated, in some future decade…. Scientific understanding of human behavior and other relevant phenomena has reached a point at which experts have identified homeostatic clusters of natural properties and mechanisms satisfying Boyd's criterion for identifying human nonmoral and moral goodness as natural kinds. Suppose also that the consequentialist part of Boyd's hunches and hopes regarding ethical theory has been vindicated…. In that case he faces the question what he should say about me (for example) if, in the face of such progress and with full knowledge of the theoretical situation, I would stubbornly refuse to assent to act-consequentialism…. What account can Boyd then give of this disagreement between us? To be precise, can Boyd give a *realistic* account of this disagreement…? Is there some procedure of empirical discovery that Boyd and I would agree would settle the issue if we could carry it out? If so, Boyd might interpret us, realistically, as debating what result that procedure would have. But I don't see what procedure Boyd could think that would be, since we are considering a situation in which all the empirical procedures Boyd believes in for deciding such matters would have settled the issue in favor of act-consequentialism…. The dispute I envisage, after all, is about act-consequentialism, and Boyd gives not the slightest reason to suppose that that could not be one of the issues in empirically irresoluble dispute between conflicting ethical systems…. A view that is forced to be nonrealist about a fundamental dispute between consequentialists and nonconsequentialists is forced to be nonrealist on an important topic, and hence in an important way, I should think. This will be an objection to virtually any naturalistic identification of ethical properties.

Boyd has written a paper in response to this argument. If the reader wishes to know what Boyd's response to this objection is, I refer the reader to that paper. What I want to focus on here is whether Adams' objection poses any problem for the version of NMR that I have been discussing. I do not think that it does. Notice that Adams says "This will be an objection to virtually any naturalistic identification of ethical properties."

I take Adams to be suggesting some case like this: On their death beds, Adams and Boyd's heads are severed and cryogenically frozen. In the distant future, their heads are thawed and placed into life supporting vats. Scientists of the distant future then inform Adams and Boyd's heads that all the empirical results predicted by Boyd's metaethical theory have come to pass. CTR, combined with empirical results, delivers the judgment that 'wrong' refers to the natural kind Boyd says it does. Boyd's head accepts CTR together with TRUTH and IDENTITY. Together these claims imply that some form of consequentialism is true. Adams' head agrees that CTR, combined with empirical results, delivers the judgment that 'wrong' refers to the property of failing to maximize utility. Adams' head accepts the relevant empirical claims. But he doesn't accept the speculative philosophical claims about CTR, TRUTH, and IDENTITY that Boyd's head does. Maybe Adams' head accepts CTR so that he thinks 'wrong' really does refer to the relevant natural kinds. But he doesn't buy into TRUTH or IDENTITY. Maybe he thinks that CTR is nothing more than a boring semantic theory and that questions about reference do not have the incredibly interesting and important consequences that proponents of TRUTH and IDENTITY attach to them. Maybe instead Adams' head does not accept CTR. Maybe Adams' head accepts some other theory of reference—perhaps some form of descriptivism—and he thinks that form of descriptivism plus TRUTH and IDENTITY together deliver the judgment that consequentialism is false. Or, finally, maybe Adams' head just doesn't buy into any of the theories or theses I have associated with NMR. Maybe he accepts all the relevant empirical claims but he thinks that CTR, TRUTH and IDENTITY are all false. So I think that the proponent of NMR can easily give a realistic interpretation of Adams' head's disagreement with Boyd's head. Adams' head accepts all of the relevant empirical claims. That isn't what the disagreement is about. The disagreement is instead located in Adams' head's rejection of some or all of the speculative philosophical and semantic theories and theses that proponents of NMR add to the relevant empirical hypotheses to construct their metaethical theories. Granted, sometimes proponents of NMR write as though they think that philosophy departments should close down and all the relevant philosophical work should be done by scientists working in labs and looking through microscopes. So insofar as Adams' objection is an expression of puzzlement about that claim, then I share his puzzlement. But when proponents of NMR write that way they are exaggerating. They don't really mean it. They employ substantive philosophical theories and theses such as CTR, TRUTH, and IDENTITY to get their metaethical theories out of their empirical claims. As a result, proponents of NMR, Boyd included, can easily give an account of the disagreement between Adams' head and Boyd's head.

It is possible that Adams is suggesting a different sort of case. Maybe the idea is this: Adams and Boyd's heads are frozen and then thawed in the distant future just as before. Again the scientists come in and tell Adams and Boyd's heads that all of the empirical predictions made by Boyd's metaethical theory are true. But then the philosophers of the distant future come in and inform Adams and Boyd's heads that all philosophers now accept CTR, TRUTH, and IDENTITY and they all agree that together with the empirical data these theories and theses deliver the result that Boyd's homeostatic consequentialism is true. The philosophers share with Adams and Boyd's heads all the relevant arguments that have persuaded future philosophers of these things. Adams' head then assents to all of this. He agrees that CTR, TRUTH, and IDENTITY are all true, he agrees that the empirical predictions made by Boyd have come true, and he agrees that together these things imply that Boyd’s homeostatic consequentialism is true. But Adams’ head still denies that homeostatic consequentialism is true. If this is the case Adams means to be suggesting, I think Boyd can still give a realistic account of the disagreement. Adams’ head believes a bunch of philosophical and scientific things that leave no space for any metaethical theory other than Boyd’s homeostatic consequentialism. But Adams’ head strangely rejects homeostatic consequentialism anyway. His belief is somehow isolated from all his other philosophical and empirical beliefs. So there is disagreement here. Adams and Boyd’s heads clearly believe different things.

Maybe Adams’ claim is not that NMR can’t give an account of Adams’ and Boyd’s heads disagreement with respect to belief in the proposition that NMR is true. Maybe instead his complaint is that NMR can’t give an account of rational disagreement. The proponent of NMR must say that Adams’ head is being irrational. If that is Adams’ complaint, then I don’t think it is very persuasive. Going against my usual anti-naturalist inclinations I’d have to side with the proponents of NMR on this one. Adams’ head, if he adopts all the relevant philosophical theories and theses together with the relevant empirical claims and he continues to think that these things imply Boyd’s consequentialism but he still doesn’t believe it, is pretty clearly irrational.

There is in fact a very interesting argument from disagreement against NMR that, like Adams objection, seeks inspiration from Moore’s Open Question Argument. It is the Moral Twin Earth objection due to Horgan and Timmons. There are some similarities between Adams’ objection and the Moral Twin Earth objection. So one might reasonably think that Adams is gesturing towards a Moral Twin Earth type of objection. In that case he would certainly be onto an interesting problem for the theory. But if that is what Adams is gesturing towards, then once again the objection is unoriginal. The widely discussed Moral Twin Earth objection to NMR appeared in the literature at least a decade before *Finite and Infinite Goods* and, as with the Qua Problem, there are many responses by proponents of NMR and counter responses by those skeptical of NMR. So there might be a problem here for NMR. But Adams' objection doesn't give the proponents of NMR anything new to worry about.

2.5 Rea's Objection

Another interesting objection to NMR due to Michael Rea. It seems to me that Rea's paper is naturally interpreted as arguing against a variety of different formulations of NMR. I think that many of these arguments are very interesting and persuasive. But what I want to focus on here is the section of Rea's paper that argues against the sort of NMR that I have been discussing. As far as I can tell, Rea argues against the version of NMR that interests me in just one short passage:

> One might think that we could go some distance toward showing that a particular reduction is true if we could show that the reduction in question has correctly identified nonmoral properties (or clusters of properties) that are tracked by our actual use of the terms 'morally good' and 'morally right'. But even if we could show this, we would still not have enough to show how belief in objective moral properties is justified. Consider the following two premises:
> (1) If there are objective moral properties, and if theory T of the nature of morality, rationality, and related notions is correct, then moral properties are identical with or composed of natural properties N1 – Nn.
> (2) Our uses of words that allegedly refer to moral properties reliably track N1 – Nn.
> Perhaps some interesting conclusions follow from these premises. But clearly the conclusion that there are objective moral properties does *not* follow from the premises. Thus,… even if it can be shown that a particular reduction has correctly identified natural properties tracked by our moral terms, there is still work for a naturalist to do in showing how belief in objective moral properties could be justified by the methods of science.

As I see it, there are two ways to interpret Rea's objection. One way emphasizes his talk of "alleged reference" and of a "tracking relation." This interpretation can be stated as follows: Distinguish between the reference relation and the tracking relation. Naturalists have defended the claim that moral terms in some sense “track” natural properties. Suppose they are right. Even granting this supposition, the tracking relation, whatever it may be, is no substitute for a genuine reference relation. Showing that moral terms “track” natural properties does not entail that moral terms refer to natural properties and it certainly does not entail the very ambitious claim that moral realism is true. So if the goal is to show that naturalism can accommodate moral realism (or more carefully if the goal is to show that, if empirical considerations turn out as naturalists hope, then naturalism can accommodate moral realism), then naturalists have not done anything to make that result plausible.

In assessing Rea’s argument, it is important to get clear about who his target is. In one of the footnotes to this passage, Rea makes clear that his target is Richard Boyd. Rea says that “Boyd (1988) presses this point [i.e. the point that moral terms “track” natural properties] in his own attempt to show that moral properties are reducible to natural properties.” Rea seems to be attributing (1) and (2) to Boyd. If naturalists like Boyd only appealed to a vague “tracking” relation in the way that (1) and (2) suggest, then they could be dismissed as easily as Rea dismisses them in the present passage. But many proponents of NMR, Boyd included among them, explicitly appeal to CTR. Boyd (1988), for example, says that “a naturalistic account is provided by recent causal theories of reference.” Rather than (1) and (2), the version of NMR that I am discussing, and that I think is most charitably attributable to the proponents of NMR, employs the following claims:

(1*): TRUTH and IDENTITY are true.

(2*): CTR is true and CTR delivers the judgment that moral terms refer to objective natural properties.

And as we have already seen, if causal and historical factors turn out the way naturalists hope, it is possible to employ CTR, (1*), and (2*) in defense of NMR. So it is not as though Boyd and the other naturalists under discussion here only appeal to some vague "tracking" relation between moral terms and natural properties and then mysteriously extrapolate from that the truth of NMR (though, admittedly, they sometimes write that way; so there is certainly some textual support for Rea's attribution). Proponents of NMR instead appeal to a well known theory of reference, CTR, and they supplement CTR with the (admittedly controversial) semantic and metaphysical principles mentioned in (1*). So if Boyd's hope that CTR, combined with empirical considerations about humans speakers and their environments, delivers the judgment that moral terms refer to objective natural properties is correct, then this variant of NMR cannot be dismissed as easily as Rea dismisses it. If we grant that moral terms bear the relations to natural properties that authors like Boyd say they do, then to challenge NMR, we must challenge CTR, TRUTH, IDENTITY or the defense of NMR constructed from them that I outlined above. On the present interpretation, Rea has challenged none of these things. So NMR, or at least the version of NMR I discuss here, is able to evade Rea's objection.

There is another way to interpret Rea. This way downplays his talk of "alleged reference" and of a "tracking relation." Despite his use of language like "alleged reference" maybe he really means to grant the naturalist the claim that moral terms refer to natural stuff such that *being that stuff* is an objective natural kind property. Maybe he is instead puzzled about IDENTITY's move from claims about reference relations between words and stuff to claims about one property being identical to another.

If this is really what Rea is worried about, then the force of his objection is significantly blunted. For one thing, Rea says in a footnote[33] that the objection to NMR in the quoted passage is “somewhat related” to one raised by Robert Merrihew Adams in *Finite and Infinite Goods*. However, as Adams himself emphasizes[34], his own theistic version of moral realism[35], like the version of NMR I have been discussing, relies on IDENTITY:

> I freely grant, for example, that ‘good’ does not mean the same as ‘resembles God’, though I do favor the view that the goodness of finite things consists in a sort of resemblance to God…. This approach has been developed most famously with respect to natural kinds…. [T]he property of being water is, necessarily, identical with the property of being H2O…. The causal relations between concrete samples of water, on the one hand, and users and uses of the word ‘water’, on the other hand, serve to “fix the reference” of the word—that is, to determine which stuff the word names. These ideas are commonly treated as if they were only about natural kinds…. I am proposing that we do use ethical terms in an analogous way….

So if Rea is really pressing the proponent of NMR on this move, then it is not a problem for NMR only, but also for Adams’ theistic moral realism. In fact it is not proponents of NMR like Boyd but instead Adams, in his early work on the Divine Command Theory, who was the first metaethicist to take this sort of line. Given that Rea’s overall project is not just to argue against naturalism but also to recommend theism over naturalism, this is a disappointing result for those of us who are fans of Rea’s work. In any case, since Adams clearly embraces IDENTITY and since Rea likens his objection to one due to Adams (and given that we have to downplay things Rea explicitly says), it seems doubtful that this is the right way to interpret Rea.

But suppose one does interpret Rea’s objection this way. If Rea really is just worried about IDENTITY, then it is not a specific problem for NMR or even for naturalism or even for proponents of CTR. It is a problem for a fashionable metaphysical

[33] 22

[34] 15-6

[35] In two other papers I defend Adams’ theistic version of moral realism.

and semantic picture that many contemporary philosophers embrace. Perhaps the motivation for IDENTITY is often left unexplained and it is treated simply as a platitude or an assumption. Perhaps the move in question deserves much more scrutiny and suspicion than it is usually accorded. But IDENTITY is widely embraced. So if Rea is, as I am interpreting him, suspicious of IDENTITY, then he needs to do more than just highlight it and deny it. Given how (perhaps unjustifiably) fashionable IDENTITY is, Rea needs to offer some arguments against it if his objection to NMR is going to have any force. No such arguments have, insofar as I can tell, been offered by Rea.

2. 6 Plantinga's Objection

Plantinga's criticism of NMR is in broad outline very similar to Rea's. Recall that, according to Rea, naturalists attempt to show that NMR is true by showing that moral terms "track" natural properties. As Rea points out, however, this vague tracking relation between moral properties and natural properties is not enough to secure the truth of NMR. Plantinga's line is very similar. According to him, proponents of NMR claim that moral properties are logically equivalent to natural properties. But, just as Rea's tracking relation isn't enough to secure the truth of NMR, neither, Plantinga claims, is logical equivalence. The proponent of NMR needs a "tighter" relation to secure the truth of NMR. But, this tighter relation, it is alleged, is nowhere to be found.

Plantinga's criticism of NMR, while similar in overall structure to Rea's, is more detailed and employs a few technical notions. Before discussing the precise details of Plantinga's argument, therefore, it is important to get clear about what some of these technical terms are.

Abundantism: Any and only properties that are conceptually equivalent are identical.

Sparsism: Any and only properties that are logically equivalent are identical.

Naturalistically Acceptable: A property is naturalistically acceptable if and only if that property's being instantiated does not entail the falsity of naturalism.

Equivalence: For every moral property, *m*, there is some natural property, *n*, that is such that necessarily an object instantiates *m* if and only if that object instantiates *n*.

DCT: An action is obligatory if and only if that action has the property of being such that it is an essential property of God to command all rational creatures to perform that action.

The DCT Property: The property of being such that it is an essential property of God to command all rational creatures to perform.

2.6.1 The Abundantist Equivalence Argument

Plantinga attributes to proponents of NMR the following sort of argument: Suppose, as EQUIVALENCE states, that each moral property is necessarily coinstantiated with some natural property. If this is the case, then surely naturalism can accommodate moral realism. We can say it this way:

(1) EQUIVALENCE is true.
(2) If EQUIVALENCE is true, then moral properties are naturalistically acceptable.
(3) Therefore, moral properties are naturalistically acceptable.

Plantinga does not challenge (1). In fact, much of his paper is devoted to showing that (1) is true. What Plantinga doubts is (2). He does not do much to motivate (2) on behalf of the naturalist except to say that it is an "intuitively plausible way, perhaps the most plausible way, to make a case for the thought that naturalism can accommodate morality." Despite the intuitive plausibility that he attaches to (2), however, Plantinga believes it to be false. He does not think that EQUIVALENCE carries with it the implication that moral properties are naturalistically acceptable. His reasoning is as follows: Consider DCT. According to DCT, an action is obligatory if and only if it is an essential property of God to command all rational creatures to perform that action. So, according to Plantinga, obligation is instantiated if and only if God exists. So if DCT were true, moral properties would not be naturalistically acceptable. Now, in spite of all of this EQUIVALENCE would still be true if DCT were true. Note that God is a necessary being. It is an essential property of His to command rational creatures to refrain from murder, theft, adultery, and covetousness. It is an essential property of His to command rational creatures to treat others with love and respect. These properties are natural. So there is a natural property that is necessarily coinstantiated with obligation. So if DCT were true, EQUIVALENCE would still be true. Thus, if DCT were true, it would be the case that both EQUIVALENCE is true and moral properties are not naturalistically acceptable. But if EQUIVALENCE could be true even if moral properties are naturalistically unacceptable, (2) is false. The truth of EQUIVALENCE, therefore, does not imply that moral properties are naturalistically acceptable.

2.6.2 The Sparsist Equivalence Argument

So mere EQUIVALENCE is not a "tight" enough relation between natural properties and moral properties to do the work that the proponent of NMR needs to be done. What about identity? The proponent of NMR can say that moral properties are identical to natural properties. Identity is the tightest relation available. So surely if the proponent of NMR could show that moral properties are identical to natural properties, that would be enough to show that moral properties are naturalistically acceptable. Moreover, the proponent of NMR could argue for this identity claim by adopting SPARSISM. Combing SPARSISM and EQIUVALENCE, we obtain the result that moral properties are identical to natural properties. The argument can be stated this way:

(1) Sparsism is true.
(2) Equivalence is true.
(3) If (1) and (2), then moral properties are identical to natural properties.
(4) If moral properties are identical to natural properties, then moral properties are naturalistically acceptable.
(5) Therefore, moral properties are naturalistically acceptable.

Plantinga does not think that the Sparsist Equivalence Argument is any more persuasive than the Abundantist Equivalence Argument. Premise (4), he claims, is false. Again, consider DCT. By the reasoning given earlier, if DCT were true, EQUIVALENCE would still be true. So some natural property would be logically equivalent to obligation. Given SPARSISM, therefore, obligation would be identical to some natural property. But, if DCT were true, the DCT property would be logically equivalent to obligation as well. So, again given SPARSISM, obligation would be identical to the DCT property which would itself be identical to the relevant natural property. But the DCT property is not naturalistically acceptable. So even though obligation is identical to a natural property, it is not naturalistically acceptable. Therefore, (4) is false.

2.6.3 Problems for Plantinga's Objection

I think there are some problems with the way that certain portions of Plantinga's objections are formulated. First, Plantinga doesn't do enough to show that if DCT were true, then EQUIVALENCE would be true. He says very little about why we should think EQUIVALENCE is true given DCT. All he says is this:

> What makes an action prima facie obligatory, then, would be a property that obviously entails that there is such a person as God [i.e. the DCT property]; moral obligation, therefore, would presumably be naturalistically unacceptable in excelsis. Even so, however, it would still be the case, by the above argument, that there is a descriptive property equivalent to obligation. And that property might be naturalistic as well as descriptive. For suppose, as theists typically think, God is a necessary being and it is an essential property of God to command persons to tell the truth, and to refrain from murder, theft, adultery and covetousness; more generally, suppose it is essential to God to command persons to treat others with love and respect. These properties and their complements are naturalistic; hence under these conditions there will be a naturalistic property

> equivalent to moral obligation, despite the fact that what makes an action morally obligatory obviously entails that there is such a person as God. Hence finding a naturalistic property that is logically equivalent to obligation doesn't show for a moment that obligation is itself naturalistic.

It seems to me that Plantinga's reasoning here is mistaken. Note that God could have actualized a world, W, inhabited only by immaterial souls. Since there are only immaterial souls at W, there are no instantiated natural properties. And given what Plantinga says about God's essential properties, He will command the inhabitants of W to love and respect one another and to refrain from murder, theft, adultery, and covetousness. By DCT, therefore, the inhabitants of W incur obligations. So obligation is instantiated at W. But there are no instantiated natural properties at W. So obligation isn't coinstantiated with any natural properties at W. So obligation isn't *necessarily* coinstantiated with a natural property. So if Plantinga's version of DCT were true, EQUIVALENCE would be false. Of course, Plantinga's formulation of DCT delivers an EQUIVALENCE like result—that moral properties are necessarily coinstantiated with non-moral or factual or descriptive properties—in this case theological properties. But theological properties are not natural. So Plantinga's DCT does not get him EQUIVALENCE which is what he needs to respond to the arguments from Equivalence.

It seems to me that focusing so myopically on the details of a particular version of DCT makes Plantinga's objection unnecessarily complicated. Just pick some non-naturalist metaethical theory that is incompatible with naturalism, say some form of Mooreanism. If Mooreanism is true, then naturalism is false. The naturalist's austere ontology does not allow Moorean non-natural moral properties or anything like that. So a theory that posits such entities entails the falsity of naturalism. But, according to Mooreanism, EQUIVALENCE is true. So the proponent of NMR can't really claim to advance the debate by showing that EQUIVALENCE is true. After all, EQUIVALENCE is entailed by Mooreanism. So how could the truth of EQUIVALENCE help show that moral properties are naturalistically acceptable when EQUIVALENCE is assumed by parties to the debate that reject naturalism? So I think my complaints above are mere formulational quibbles about Plantinga's use of DCT rather than substantive problems for his overall point.

Second, although Plantinga uses phrases such as 'If DCT were true, then ____' he is not clear about whether such phrases are supposed to be understood as counterfactuals or counter possibles. Furthermore, the text contains evidence that it is neither. Note that Plantinga believes all counterpossibles are trivially true:

> [I]t is an essential property of God not to command hate instead of love. There aren't any possible worlds in which God commands hate rather than love. True, at least on the usual semantics for counterfactuals, if God had commanded hate, hate would have been right and love wrong. But this is of no more interest than the fact that if there were no prime numbers, all numbers would be prime.

Given what he says it is therefore doubtful that the relevant phrases are meant to be counterpossibles. So are they counterfactuals meant to denote metaphysical possibility? I doubt it. Consider the way that Plantinga sets up his investigation into whether naturalism can accommodate moral realism. He starts by saying this:

> What, exactly, or even approximately, is this 'accommodating'? We might begin by returning to the question I raised above: is
> (1) If naturalism were true there would be no such thing as moral obligation— that is, no actions would possess the property of being morally obligatory
> true? Here we immediately run into serious problems having to do with the modal status of theism and naturalism…. This way of thinking about our question—can naturalism accommodate moral obligation?—runs into a thicket of difficulties, difficulties arising from the noncontingent nature of theism, and perhaps also the noncontingent nature of propositions involving the existence of moral obligation. Such difficulties are familiar, certainly; but that doesn't make them any more tractable. This isn't the place to try to figure out how to reason about noncontingent propositions of this sort; that would require more than a whole paper on its own account. What is clear, however, is that addressing our question by way of asking after the truth of (1) does not promise to be fruitful.

These issues are what motivate Plantinga to discuss the question of whether naturalism can accommodate moral realism in terms of whether the arguments from Equivalence are sound rather than in terms of counterfactuals like (1). More than this, the same difficulties Plantinga cites in his discussion of (1) spring up with respect to ‘If DCT were true, then ____’ when construed as a counterfactual. Imagine that you are a naturalist and you hear Plantinga say “If DCT were true, then ____”. You might reasonably respond as follows: Is this statement a counterfactual or a counterpossible? It can’t be a counterpossible because those have trivial and uninteresting truth-values and are not to be used in evaluating theories. But how could it be a counterfactual? As a naturalist, you will think that God doesn’t exist. And, as Plantinga points out, whether God exists is non-contingent. If He exists He exists necessarily. If He doesn’t exist, then He doesn’t exist at any possible world at all. So the antecedent of the phrase ‘If DCT were true, then ___’ is necessarily false. So if you are right about naturalism, then the relevant phrase is a counterpossible rather than a counterfactual. But if Plantinga and I are right about God’s existence, then the relevant phrase is a counterfactual. Since a counterfactual is what Plantinga needs to make the argument work and since the relevant phrase is a counterfactual only if God exists, then the success of Plantinga’s objection depends on the existence of God. Why should you, the naturalist, be bothered by that?

I think this is a mere formulational quibble and that Plantinga has an interesting point. For one thing, I am open to holding that counterpossibles have substantive non-trivial truth-values (although a robust naturalist may not be in which case we run into the same problems, such a naturalist might not even think that farfetched counterfactuals should be used in evaluating theories). For another thing, I think that Plantinga can make his point without employing a counterfactual or a counterpossible. He could just say something like this: Some people claim that moral properties are naturalistically acceptable. They try to show this by showing that EQUIVALENCE is true. But according to other moral theories quite hostile to naturalism, EQUIVALENCE is true. So how does just showing that EQUIVALENCE is true show that moral properties are naturalistically acceptable? The naturalist has much more work to do to establish that. I think this is a fair point to make and I take this to be the overall drift of Plantinga's argument. And perhaps, as Plantinga suggests, naturalists sometimes write as if they are endorsing one of the arguments from Equivalence. So insofar as they write that way, I am sympathetic to Plantinga's point. But it is important to get clear about what Plantinga thinks his arguments have accomplished and who his targets are. First, consider what Plantinga suggests he has established:

> It looks as if, if abundantism is true, there is no way to argue cogently that obligation is naturalistically acceptable. On the other hand, if sparsism is true, then not even showing that obligation is identical with some naturalistic property will suffice to show that obligation is naturalistically acceptable; for obligation might well be identical with a naturalistic property, but also identical with a property obviously entailing that there is such a person as God. It therefore looks as if there is no way at all of cogently arguing that naturalism can accommodate moral obligation…. By way of conclusion: [my objection] presents a real (I would say insoluble) problem for one who wants to make a case for the idea that metaphysical naturalism can accommodate morality.

Plantinga seems to be suggesting in the passage above that Sparsism and Abundantism exhaust all possibilities for the proponent of NMR. If the naturalist wants to show that NMR is true, then her last best hope is to employ one of the Arguments from Equivalence Plantinga has criticized. If, therefore, Plantinga's criticisms of these arguments are sound, then the proponent of NMR is left without any way of arguing that naturalism can accommodate moral realism. More specifically, note that one of Plantinga's targets is Cornell Realism. He says this:

> "Cornell realists"… have maintained that naturalism can perfectly well accommodate the existence and exemplification of specifically moral properties, including moral obligation. What I want to investigate is this question: is it true that naturalism, taken as above, can accommodate morality?

> Now the first question is whether [the proponent of NMR] takes [a natural property] to be equivalent to rightness but distinct from it, or whether he takes it to be identical with rightness. According to Jackson, "Cornell realists" take it that "ethical properties are identical with descriptive properties"… if he's right, perhaps [the proponent of NMR] holds that [the relevant natural property] is identical with rightness…. Suppose, then, that sparsism is true. That means, of course, that rightness is indeed identical with P. As we've seen, however, this, even if true, doesn't at all show that rightness is naturalistically acceptable.

In this passage Plantinga attributes SPARSISM to the Cornell Realists. His reasoning seems to be that the Cornell Realists believe that moral properties are identical to natural properties. SPARSISM explains that. So the Cornell Realists must accept SPARSISM.

I think that Plantinga is mistaken about the scope of his criticism and mistaken in his attribution of SPARSISM to the Cornell Realists. I believe that the version of NMR that I have been discussing throughout this paper is true to the spirit of Cornell Realism. So let us attribute to Cornell Realists CTR, IDENTITY, and TRUTH. It is possible, if one accepts these theories and theses, to hold that moral properties are identical to natural properties without endorsing SPARSISM. So the proponent of NMR might think that obligation is identical to some natural property and at the same time avoid commitment to the claim that all logically equivalent properties are identical properties. As I will show below, CTR, together with IDENTITY and TRUTH, allow the naturalist maintain just this claim. And, as we will see, without the attribution of SPARSISM, Plantinga's objection to the version of NMR that I have been discussing is unsuccessful.

So suppose the proponent of NMR adopts CTR, IDENTITY, and TRUTH. Then whether NMR is true will depend on whether CTR delivers the judgment that moral terms refer to objective natural properties. Whether CTR delivers this judgment, the proponent of NMR will claim, is an empirical question. Suppose we grant that CTR does deliver that judgment. Suppose in particular that 'obligation' refers to some natural property P. By IDENTITY, obligation and P are identical properties. So the proponent of CTR might think that establishing that ‘obligation’ refers to a natural objective property P will, combined with IDENTITY, yield the result that moral realism about obligation is true. Suppose that some similar story can be told for the other moral terms. They each refer to a natural objective property. Then NMR is true.

So how could Plantinga’s strategy for objecting to the Equivalence Arguments be extended to the version of NMR presently under discussion? Well, recall that the dialectic with respect to the Equivalence Arguments went as follows. The naturalist identifies some relation, in particular EQUIVALENCE, between moral properties and natural properties. The naturalist then claims that if the relevant relation holds, then NMR is true. Plantinga replies by arguing that even if DCT were true, the relevant relation would hold between moral and natural properties. But NMR would be false. So moral properties would still be naturalistically unacceptable.

Perhaps, then, Plantinga’s response could be extended to the version of NMR I am discussing in the following way: The naturalist identifies a relation, in this case the reference relation between moral terms and natural properties rather than an equivalence relation between moral properties and natural properties. CTR judges that moral terms refer to natural properties. Combined with TRUTH and IDENTITY it is alleged that the relevant reference relation establishes the truth of NMR. Now, Plantinga might claim that if DCT were true, then moral terms still refer to natural properties and that TRUTH and IDENTITY would still be true. So if DCT were true, moral terms would still refer to natural properties. So just showing that moral properties refer to natural properties does not show that they are naturalistically acceptable.

But clearly this strategy will not work. If DCT were true and CTR, TRUTH, and IDENTITY were all true, then some story like the following would have to be true: God issues some commands and instructs Moses to tell people about those commands. Moses sees someone acting in accordance with one of those commands. He points to the relevant action and utters some sentence like 'See what he did! That is obligatory. All of us should do that too.' Now, given CTR's account of reference grounding, something about Moses and his environment, together with his act of pointing and uttering determine that 'obligation' refers to the DCT property. Of course, the relevant act will be a sample of many other things besides the DCT property. It will be a sample of maximizing the world's hedonic index. It will be a sample of physically moving around in some precise way. And it will be a sample of numerous other things. One therefore might wonder why it is that CTR in this case delivers the judgment that 'obligation' here refers to the DCT property rather than any of these other things. But if that is the worry then it is just the Qua Problem. So if Plantinga's objection is supposed to be independent of the Qua Problem and something new for the proponent of NMR to worry about, then we have to grant that the referent of 'obligation' in this case is fixed to accordance with God's commands rather than to anything else. But the DCT property is not a natural property. So if DCT were true, it would not be the case that 'obligation' referred to some natural property. And since, the Qua Problem notwithstanding, moral terms do not in the present case refer to any natural properties, TRUTH and IDENTITY do not cause the same trouble for the version of NMR I have been discussing that SPARSISM allegedly caused for the proponent of the Sparsist Equivalence Argument. Thus, Plantinga's objections to the Equivalence Arguments cannot be extended to the version of NMR that I have been discussing without making them dependent upon the Qua Problem.

2.7 Conclusion

Robert Adams, Alvin Plantinga and Michael Rea have defended some objections to NMR. I have argued that each of their objections is either unpersuasive or dependent upon objections already widely discussed in the literature on NMR.

CHAPTER 3

MUST EARTH AND TWIN EARTH DIVERGE?

David Brink, Neil Levy, and David Merli have defended Naturalist Moral Realism (NMR) from Horgan and Timmons' Moral Twin Earth thought experiment. Although they make their points in different ways, each of their defenses relies on the same basic idea: On any consistent interpretation of the thought experiment, there is some important respect in which Earth and Twin Earth will eventually diverge. Once this respect of divergence is appreciated, it is alleged, the thought experiment no longer threatens NMR. In this paper I argue that Brink, Levy, and Merli are mistaken. Earth and Twin Earth need not diverge in the way that these authors suggest. And even if such divergence were certain to occur, it would not in any way undermine the force of Horgan and Timmons' thought experiment.

3.1 Naturalist Moral Realism

To begin with, it will be useful to briefly review some of the terms of art, and some of the assumptions, found in the literature on Moral Twin Earth:

Causal Regulation: Causal regulation[36] is a relation that obtains between a set of properties and a term. A set of properties causally regulates a term if and only if "there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term… will be approximately true of" the set of properties.

Reference: A moral term refers to a set of properties if and only if that set of properties uniquely causally regulates that moral term.

Tc and Td: Tc is a consequentialist theory. Td is a deontological theory.

[36] Causal regulation is introduced in Boyd (1988).

Tc and Td Properties: Tc-properties is a set of properties whose essence is captured by Tc. Td-properties is a set of properties whose essence is captured by Td.

NMR: NMR is naturalist moral realism. NMR holds that 'right' on Earth is uniquely causally regulated by Tc-properties, that *Reference* is the correct theory of reference for moral terms, that these claims about causal regulation and reference provide the resources to establish that there are objective moral properties, and that such properties are natural and in some sense "scientifically respectable."

Twin Earth: Twin Earth is a near duplicate of Earth somewhere in our galaxy. It is only different from Earth in that on Twin Earth 'right' is uniquely causally regulated by Td-properties rather than Tc-properties and any other differences entailed by this one difference.

3.2 The Moral Twin Earth Argument

With these terms of art in hand, we are now in a position to state Horgan and Timmons' Moral Twin Earth thought experiment. Consider the following case: Some of the inhabitants of Earth meet some of the inhabitants of Twin Earth. Together they witness an organ harvest. Two people were dying. One needed a new heart. The other needed a new pair of lungs. A third person, who was completely healthy, was killed so that his organs could be used to save the lives of the other two. The organ harvest maximized utility. So Tc dictates that it was right. However, the organ harvest required killing one to save the lives of two. This is a violation of one of Td's rules. So Td dictates that it was not right. The inhabitants of Earth say 'The organ harvest was right'. The inhabitants of Twin Earth say 'The organ harvest was not right'. The inhabitants of both planets believe that they are disagreeing with each other. Each side offers the other reasons for why they should adopt their respective claims. Neither side changes its mind.

This case motivates an argument against NMR. If NMR is true, then the inhabitants of Earth are not disagreeing with the inhabitants of Twin Earth. They are just talking past each other. When the inhabitants of Earth use ‘right’ they do not refer to the same set of properties that the inhabitants of Twin Earth refer to when they use ‘right’. So they are not disagreeing with each other. But this is a crazy result. The inhabitants of Earth are, in fact, disagreeing with the inhabitants of Twin Earth. They are not just talking past each other. So NMR is not true.

3.3 Brink on Diverging Worlds

My goal in this paper is to close off one particular response to this argument. The response that interests me here has its origins in a paper by David Brink. Brink grants, at least for the sake of argument, that Horgan and Timmons’ thought experiment refutes its intended target—extant formulations of NMR. Brink then proceeds to modify NMR by changing the account of reference. It is his hope that this change will allow NMR to deal with Horgan and Timmons’ objection. As far as I can tell, Brink’s modification of NMR has not been widely adopted. However, a different response, briefly suggested by Brink in a footnote, has made more of an impact. In the footnote, Brink says the following:

> If people have the same commitments to morality on Earth and Moral Twin Earth, the differing standards will cause each planet’s people to assess people, actions, and institutions differently; over the long run, this should affect the course of individual and social histories on Earth and Moral Twin Earth. Though the members of both planetary pairs—Earth and Twin Earth and Earth and Moral Twin Earth—are…otherwise indistinguishable, this caveat includes many more differences in the second pair than in the first. As it seemed important to Putnam’s original arguments that differences between Earth and Twin Earth be minimized, the more extensive differences between Earth and Moral Twin Earth may complicate Timmons and Horgan’s argument (2001: 165, n21).

This idea of Brink's, that on any coherent interpretation of the thought experiment Earth and Twin Earth will diverge in a way that might serve to undermine the intuition doing the work in Horgan and Timmons' objection, has been adopted and developed in detail by Neil Levy and David Merli[37]. In what follows, I will consider those developments and conclude that none of them is successful. Brink's idea, I will argue, cannot be made to work. The proponent of NMR must find some other way to respond to Horgan and Timmons' thought experiment.

3.4 The Argument from Twin Earth Psychology

Horgan and Timmons stipulate that 'right' on Earth is uniquely causally regulated by Tc-properties while 'right' on Twin Earth is uniquely causally regulated by Td-properties. They claim that this difference between Earth and Twin Earth can be explained by psychological differences between the inhabitants of Earth and the inhabitants of Twin Earth. The inhabitants of Earth are prone to sympathy and not to guilt. The inhabitants of Twin Earth are prone to guilt and not to sympathy. This difference, it is alleged, explains why 'right' is causally regulated by different properties on different planets. Horgan and Timmons make the point this way:

> The differences in causal regulation, we may suppose, are due to species-wide differences in psychological temperament that distinguish Twin Earthlings from Earthlings. (For instance, perhaps Twin Earthlings tend to experience the sentiment of guilt more readily and more intensively, and tend to experience sympathy less readily and less intensively, than do Earthlings)

Perhaps the explanation Horgan and Timmons have in mind is this: If the inhabitants of Twin Earth suffer a lot of guilt when they violate Td's rules and only a little compassion, then they will form the belief that adhering to Td's rules is right when Td's rules conflict with maximizing utility. On the other hand, if the inhabitants of Earth feel a lot of

[37] For another interpretation of Brink's footnote, see Rubin (2008).

compassion and only a little guilt when they violate one of Td's rules, then they will form the belief that adhering to Td's rules is not right when Td's rules conflict with maximizing utility. In the former case, what is predicated of 'right' will more often be true of Td-properties than Tc-properties. In the latter case, what is predicated of 'right' will more often be true of Tc-properties than Td-properties. So, other things being equal, in the former case Td-properties will uniquely causally regulate 'right' and in the latter case Tc-properties will uniquely causally regulate 'right'.

One of the ways in which Neil Levy has developed Brink's suggestion is by arguing that this stipulation about the psychological differences between the inhabitants of Earth and Twin Earth renders Horgan and Timmons' thought experiment inconsistent. Rather than explaining why 'right' on Twin Earth is causally regulated by Td-properties, the stipulated psychological differences between the inhabitants of Earth and Twin Earth instead imply that 'right' on Twin Earth is causally regulated by Tc-properties just as it is on Earth. Levy (forthcoming, pp. 5-7) makes the point this way:

> In assessing Horgan and Timmons' claim that on [Twin Earth] moral terms are causally regulated by a set of properties captured by some deontological theory, we need to focus on what makes this stipulation true; *why* are moral terms regulated in this way? Horgan and Timmons provide us with an answer in their thought experiment: moral terms are regulated by a deontological property on [Twin Earth] because the inhabitants differ from us psychologically. It is because they are more prone to guilt and less to sympathy that a deontological theory is true at their world. Now, how are we to interpret this stipulation…? [T]he ways in which the inhabitants of [Twin Earth] differ from us psychologically is relevant to the circumstances in which we and they act, and not to the moral theory true at our worlds. On this view, it might be (for instance) that a deontological theory is true only at a superficial level on [Twin Earth]—it is because adherence to strict rules maximizes (say) happiness at their world that they are (superficially) deontologists…. [T]he inhabitants of [Twin Earth] are only superficially deontologists. Perhaps their psychologies cause them to feel anxiety at the absence or violation of rules governing their lives, on the one hand, and ensure that they are less motivated by sympathy on the other (perhaps they feel less compassion for the distress of others who are hurt by their rules or whose distress could be alleviated only at the cost of violation of their rules). These differences from us explain why an apparently deontological theory is true [on Twin Earth]: it is because such a theory maximizes the average happiness of its inhabitants. Because they feel less sympathy, they do not feel dissatisfaction at departures from what we regard as optimal states of affairs and are less unhappy at such departures; because they are motivated to abide by rules and feel anxious at their violation, rule following is satisfying for them. But if it is the case that guiding their behavior by reference to a deontological theory maximizes their happiness, the deep justification for guiding their behaviour in that manner is, in

> fact, consequentialist. They are, in fact, fundamentally consequentialists, not deontologists at all; it is just that the ways in which they fetishize rules causes it to be true that they best achieve their consequentialist ends by rule-following…. It is therefore not the case that the inhabitants of [Twin Earth] disagree with us in their conflicting utterances, but nor is it the case that they fail to make moral claims at all. They fail to disagree with us because, properly understood, they *agree* with us.

As I interpret him, Levy's argument runs as follows: There is only one possibility that is consistent with Horgan and Timmons' alleged explanation of why 'right' on Twin Earth is causally regulated by Td-properties rather than Tc-properties. The inhabitants of Twin Earth feel anxiety at the absence or violation of the rules dictated by Td. Furthermore, the inhabitants of Twin Earth feel little compassion for the distress of those hurt by such rules and those whose distress could be alleviated by violating such rules. Given these features of Twin Earth psychology, adhering to Td's rules always maximizes utility on Twin Earth. The inhabitants of Twin Earth are so averse to violating Td's rules and feel so little compassion at the suffering of those who would benefit from a violation of Td's rules that the utility that might be achieved by violating Td's rules on Twin Earth is always outweighed by the disutility of violating Td's rules. But if adhering to Td's rules always maximizes utility on Twin Earth, then 'right' on Twin Earth is causally regulated by Tc-properties rather than Td-properties. But this contradicts Horgan and Timmons' stipulation that 'right' on Twin Earth is causally regulated by Td-properties. So, this interpretation of the thought experiment is inconsistent. The proponent of NMR need not worry about it.

3.5 Two Problems for the Argument from Twin Earth Psychology

There are two problems for Levy's argument. First, Levy claims that if Horgan and Timmons stipulation about the psychologies of the inhabitants of Twin Earth is true, then adhering to Td's rules always maximizes utility on Twin Earth. But even if Levy is right about this, it does not follow that ‘right’ on Twin Earth is causally regulated by Tc-properties. Second, Levy never motivates the claim in question. He just asserts it. And it seems easy to construct counterexamples to it.

3.5.1 ‘Right’ is not Causally Regulated by Tc-Properties Rather than Td-Properties

Concerning the first problem: There is nothing about the possibility Levy highlights that determines that ‘right’ is causally regulated by Tc-properties rather than Td-properties or both. Levy just assumes that if Tc and Td prescribe the same actions over the course of the entire history of Twin Earth, then that is enough to get the result that ‘right’ is causally regulated by Tc-properties and not Td-properties. But, for all Levy has said, both Tc-properties and Td-properties causally regulate ‘right’ in the variant of Twin Earth that Levy describes. Consider again the account of causal regulation discussed in this literature: A set of properties causally regulates a term if and only if “there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term… will be approximately true of” the set of properties. Given this account, it is true that on Levy’s variant of Twin Earth Tc-properties causally regulate ‘right’. But Td-properties also causally regulate ‘right’. The only feature of Levy’s variant that is relevant to the account of causal regulation is this: Every action on Twin Earth that is prescribed by Td is also prescribed by Tc. So the inhabitants of Twin Earth witness an action. If that action adheres to what Td prescribes then they judge that action to be right. If that action adheres to what Tc prescribes, then they judge that action to be right. If that action departs from either what Td prescribes or what Tc prescribes, then they judge that action to be wrong. So given the details of Levy’s variant on Twin Earth supplied so far, what is predicated of ‘right’ is true of Td-properties to exactly the degree to which what is predicated of ‘right’ is true of Tc-properties. So if all we have to go on are the details Levy supplies, then in every case what is predicated of ‘right’ is just as true of Td-properties as it is of Tc-properties. It would therefore seem that ‘right’ on Twin Earth is causally regulated by both Td-properties and Tc-properties. So there is no set of properties that uniquely causally regulates ‘right’ on Twin Earth. So, given *Reference*, ‘right’ on Twin Earth does not refer to anything. And if ‘right’ on Earth refers to Tc-properties and ‘right’ on Twin Earth doesn’t refer to anything, then we get the result that Horgan and Timmons want—moral disagreement between the inhabitants of Earth and the inhabitants of Twin Earth is not genuine.

If this case is going to do the work Levy needs it to do, then, it is going to have to be true that on any way of filling the out the details, ‘right’ on Twin Earth is causally regulated by Tc-properties but not Td-properties. The problem is that there are ways of filling out this case so that this is not true.

One thing missing from Levy’s variant on Twin Earth are the details concerning the inhabitants of Twin Earth’s counterfactual judgments about cases in which it doesn’t turn out that actions that adhere to Td’s rules maximize utility. Sure, as it so happens, on Twin Earth following Td’s rules always maximizes utility. But one could ask the inhabitants of Twin Earth what they think about cases, remote from their own, in which following Td’s rules doesn’t maximize utility. One could ask them to consider creatures psychologically different from them—like the inhabitants of Earth. Should the inhabitants of Earth adhere to Td’s rules? Or, one could bring to their attention a scenario, that has somehow never happened on Twin Earth, in which one of the inhabitants of Twin Earth finds herself in a situation in which, as much disutility as is produced by all the guilt and anxiety that would stem from violating one of Td's rules, the consequences of adhering to Td's rules would involve much more disutility[38]. What should an inhabitant of Twin Earth do in a situation like that?

From here the range of possibilities Levy has highlighted can be divided into three cases. In one case, the inhabitants of Twin Earth judge that Td's rules should be violated. In another case, the inhabitants of Twin Earth judge that Td's rules should be upheld and utility should not be maximized. In a third case, the inhabitants of Twin Earth just remain agnostic and refrain from making any judgments about the rightness of the relevant counterfactual actions. For exactly the reasons Levy discusses, the first of these cases is not a problem for NMR. The second and third cases, however, are a problem for NMR.

In the second case, the inhabitants of Twin Earth judge that it is right to adhere to Td's rules in the relevant counterfactual cases while sticking with Tc and maximizing utility is wrong. So what is predicated of 'right' on Twin Earth is more often true of Td-properties than Tc-properties. So if 'right' on Twin Earth is uniquely causally regulated by anything it is Td-properties rather than Tc-properties. And we therefore have a variant of Levy's case on which Horgan and Timmons' argument goes through. 'Right' on Earth refers to Tc-properties, 'right' on Twin Earth refers to Td-properties, and moral disagreement between the inhabitants of Earth and Twin Earth is not genuine.

Now consider the third case. Suppose one fills out Levy's variant on Twin Earth by stipulating that the inhabitants of Twin Earth would just be agnostic about what is right in such counterfactual situations. Then we are back to the problem that motivated dividing cases. On Twin Earth what is predicated of 'right' is true of Td-properties to

[38] See below for a more detailed discussion of such a situation.

exactly the degree to which it is true of Tc-properties. So if what is predicated of ‘right’ is approximately true of Tc-properties, then it is approximately true of Td-properties as well. So both sets of properties causally regulate ‘right’ on Twin Earth. And if what is predicated of ‘right’ is not approximately true of either set of properties, then nothing causally regulates ‘right’ on Twin Earth. So, either way, the account of reference taken for granted in this literature implies that ‘right’ on Twin Earth doesn’t refer to anything. Again, we get the result that moral disagreement between the inhabitants of Earth and Twin Earth is not genuine.

3.5.2 Following Td’s Rules Does not Always Maximize Utility on Twin Earth

Concerning the second problem: Levy’s argument relies on the following premise: If the inhabitants of Twin Earth have the psychologies stipulated by Horgan and Timmons, then following Td’s rules will always maximize utility on Twin Earth. But Levy never says anything to motivate this premise. Why think that, since the anxiety and lack of compassion are enough to cause the inhabitants of Twin Earth to judge that acting in accordance with Td’s rules is always right, acting in accordance with Td’s rules always maximizes utility on Twin Earth?

Consider the following case: Hank is hiding Jews in his attic. The Twin Earth Nazis come to Hank’s door and ask “Are you hiding any Jews?” Hank has two alternatives. He can say “No” or he can say “Yes.” If Hank says “No” he will be lying and therefore violating one of Td’s rules. This will cause him to suffer great guilt and anxiety. If Hank says “Yes” he will not violate any of Td’s rules. Given that Hank is only a little compassionate, this will cause him only a little sadness. If Hank says “No” the Twin Earth Nazis will leave and the Jews will be safe. If Hank says “Yes” the Twin

Earth Nazis will send the Jews to a concentration camp. Saying "No" violates one of Td's rules but maximizes utility. Saying "Yes" doesn't violate any of Td's rules but fails to maximize utility. Nothing about Horgan and Timmons' stipulation precludes things like this from happening on Twin Earth. Or, at the very least, Levy hasn't said why he thinks such things couldn't happen on Twin Earth.

3.6 The Argument from Twin Earth Moral Institutions and Practices

Levy and Merli have offered another argument based on Brink's suggestion. The basic idea is that the moral institutions and practices of Earth and Twin Earth will, at least in the future, be radically different. These differences are so radical that they undermine the intuition that apparent disagreement between the inhabitants of Earth and Twin Earth is genuine. And without this intuition, Horgan and Timmons' argument against NMR has no force. Levy (forthcoming, pp. 7-8) states the argument this way:

> As David Brink (2001) has pointed out, the histories of these two planets ought to diverge over time. The psychological differences alone between us and the inhabitants of [Twin Earth] would be sufficient to force a divergence, which would set the planets off on quite different trajectories; when we add to that the fact that these psychological differences entail differences in moral institutions and practices, the divergence becomes quite radical. So were we to encounter [Twin Earth] later in its (and our history), the differences would be so striking, I think it is fair to claim, that there would be little temptation to think that our moral terms had the same reference…. I hold, when we bear in mind the future state of affairs in which Earth and [Twin Earth] will have diverged, we will lose the temptation to see ourselves as disagreeing with them about which actions are right.

Merli (2002, pp. 214-216) offers essentially the same argument:

> [S]uppose Horgan and Timmons want to force the realist to admit that, by his lights, twin-moral terms *do* refer to different properties…. The way to do this is by making twin-moral practice *different* from our own moralizing. It's in virtue of these differences that 'right' can refer to different properties in different places. Fleshing out the example in this way, though, weakens the univocity intuition. We think that we share the term 'right' with Moral Twin Earth because we imagine that twin-moralists use the term in roughly the same ways. Suppose we learn that this is false, because there are significant differences between our uses of the term and theirs (differences, say not just in the extensions of the terms but the kinds of reasons they give and accept, and so on). Twin moralists apply 'right' to acts we're sure are abhorrent, they offer completely different kinds of justifications for their claims, they see our reasons as irrelevant, and so on. It seems to me that then we'd either withdraw our initial judgment that we share terms with the twin-moralists, or at least offer it with less conviction.

I take Levy and Merli to be suggesting the following thought: The moral institutions and practices recommended by Tc and Td are radically different. They are so different, that if one were to compare them, one would deny that they were about the same thing. Eventually, the inhabitants of Earth and Twin Earth will figure out which moral theory is true at their respective worlds[39]. Then they will implement the recommendations of those theories. So the moral institutions and practices of Earth and Twin Earth will eventually diverge. And the intuition that apparent Earth and Twin Earth moral disagreement is genuine is undermined.

We can add force to Levy and Merli's point by filling it out with some concrete examples: First, consider John Harris' (1975) Survival Lottery. The basic idea behind the Survival Lottery can be expressed as follows: First, give everyone a number. Second, if there are two or more people who will die without a new organ, then randomly pick one of the numbers. Third, if a person's number is picked, and if that person is healthy, then kill that person and give his or her organs to the two or more people who need them. If the person is not healthy, then repeat the third step. A major selling point of Harris' Survival Lottery is supposed to be that it would maximize utility. A lot of people die from organ failure under the current system. They and their families suffer. If we adopt the Survival Lottery, the number of people who die from organ failure will drop by at least half. That means fewer deaths and fewer grieving families. So giving up the current practice and adopting the Survival Lottery would maximize utility.

Now, Tc dictates that we maximize utility. So Tc dictates that we adopt the Survival Lottery. Td, on the other hand, dictates that one should never kill even to save a

[39] Why believe this? I don't know.

life. So Td dictates that we do not adopt the Survival Lottery. If this line of reasoning about the Survival Lottery is correct, then we can add some support for Brink, Levy, and Merli's suggestion that the moral institutions and practices of Earth and Twin Earth will radically diverge. On Earth we will eventually adopt the Survival Lottery. On Twin Earth, they will never adopt the Survival Lottery. That is a radical difference.

Next, consider the practice of voting. One might think that utilitarianism suggests that ordinary individuals have no obligation to vote. An election will never hinge on just one vote. So there is no difference in utility between voting and not voting. Tc does not require that an individual vote. One of Td's rules, however, requires that individuals vote regardless of its impact on utility. On Earth, everyone will eventually stop voting because each will recognize that doing so makes no difference to utility. On Twin Earth, everyone will continue to vote because they believe they have a duty to do so regardless of whether failing to vote would also maximize utility. This is another radical difference between Earth and Twin Earth.

In light of these considerations, one might think there is some basis for Brink, Levy, and Merli's suggestion that the moral institutions and practices of Earth and Twin Earth will radically diverge.

3.7 Two Problems for the Argument from Twin Earth Moral Institutions and Practices

It seems to me that there are two problems for this argument. First, the claim that the disagreement intuition is undermined is mistaken. Second, even if the relevant claim was not mistaken, Earth and Twin Earth need not diverge in the way that Levy and Merli suggest.

3.7.1 The Disagreement Intuition is Not Undermined

First, even if one agrees that Earth and Twin Earth must diverge in the way that Levy and Merli suggest, the intuition that disagreement between Earth and Twin Earth is genuine remains. When I think about the inhabitants of Earth and Twin Earth in the distant future and I imagine the inhabitants of Earth saying ‘The Survival Lottery is right!’ and ‘Failing to vote is not wrong!’ and the inhabitants of Twin Earth saying ‘The Survival Lottery is not right!’ and ‘Failing to vote is wrong!, my intuitions remain unchanged. It is only when no concrete details of the differences in practices between Earth and Twin Earth are provided that I am even slightly tempted to think that the disagreement might not be genuine. But when you actually specify such radical differences, the intuition that the relevant disagreement is genuine remains just as strong as before. Perhaps there are other radical differences in practices that would undermine the relevant intuition. But Levy and Merli have done nothing to uncover such differences.

To hammer the point home, suppose that in the distant future, the moral institutions and practices of Earth and Twin Earth must be as radically different as one could possibly imagine. Suppose, for instance, that on Earth 90% of all tax revenues go to feeding the poor while on Twin Earth 90% of all tax revenues go towards killing the poor and building, maintaining, and increasing a pile of corpses that reaches to the sky. The inhabitants of Earth would be shocked and horrified by the use of tax revenue on Twin Earth. The inhabitants of Twin Earth would be shocked and horrified by the use of tax revenue on Earth. Suppose, in addition, that on Earth it is customary, when having guests stay the night, to treat them as a member of one’s family while on Twin Earth it is customary to kick such guests in the neck 13 times and then to ignore them. Again the inhabitants of each planet would be shocked at the customs of the other. Now, continue this by multiplying similar divergences in practice. Finally, suppose that the inhabitants of Earth and Twin Earth give radically different justification for the actions they perform. On Earth the justification is that such actions maximize utility. On Twin Earth the justification is that such actions adhere to the deontological code that Cthulhu Lord of the Underworld has given to them.

I do not know how the moral institutions of Earth and Twin Earth could be more different than this. And yet, my intuition that disagreement between the inhabitants of the two planets is genuine remains. Suppose that the inhabitants of Earth and Twin Earth meet. The inhabitants of Earth say things like ‘It is wrong for tax revenues to go to the pile of corpses!’ and ‘It is wrong to kick one’s guests in the neck 13 times!’ while the inhabitants of Twin Earth respond ‘It is not wrong for tax revenues to go to the pile of corpses!’ and ‘It is not wrong to kick one’s guests in the neck 13 times!’. It seems obvious to me that the disagreement is genuine. The inhabitants of the two planets are not talking past one another. They just have radically different theories about what is right and wrong.

My point is just this: Radical divergence in moral institutions and practices is not by itself sufficient to undermine the disagreement intuition. Brink, Levy, and Merli need to do more than just vaguely gesture towards radical differences between moral institutions and practices on Earth and Twin Earth without actually specifying in detail what those differences are. They owe us an explicit statement of the relevant differences and an explanation as to why those differences are necessitated by the differences in which sets of properties regulate which moral terms on the two planets.

3.7.2 Earth and Twin Earth Need Not Diverge

This leads to my second point. It is difficult for me to see how the formulations of Tc and Td discussed in this literature could possibly yield the degree of divergence in moral institutions and practices present in the very radical example I just discussed. In order to get Earth and Twin Earth to diverge so radically, we had to suppose that Td was a crazy deontological theory that no one actually holds. But if we stipulate that Td is a Kantian moral theory, like the idea that an action is right if and only if it is performed without treating anyone as a mere means and if we stipulate that Tc is hedonic act utilitarianism, and not some crazy theory that says we should maximize suffering, then we will not get the judgment that the moral institutions and practices of Earth and Twin Earth must be so divergent. Neither Tc nor Td will require, as a matter of general policy, devoting 90% of tax revenues to killing the poor and building a pile of corpses that reaches to the sky. Neither theory will require that everyone kick each of their guests in the neck 13 times. Neither theory will be justified by its adherents by the fact that it conforms to a code laid down by Cthulhu.

So here is the lesson from all this: the moral institutions and practices of Earth and Twin Earth will diverge no more than what Tc and Td recommend. When constructing the case, we are free to let Tc and Td be whatever consequentialist and deontological theories respectively that we want. So let us stipulate that they are the most regular and obvious and non-exotic versions of these theories. Then Tc and Td will overlap in extension far more than they do in the very radical example just discussed.

Furthermore, it is not hard to discover exactly how much the two theories fail to overlap and the exact extent to which they will recommend different moral institutions and practices. Fortunately, that hard work has already been done for us. We need only to look at the terrain already explored by normative ethicists. When we do this, we will see that, at the very most, Tc and Td diverge in what they recommend about things like the Survival Lottery and voting and similar cases. This is the limit, therefore, on how divergent Earth and Twin Earth can get. We can, therefore, rule out the idea that on Twin Earth piles of corpses reach to the sky. So this brings us back to the case of more modest divergence.

As I said above, my intuition about the more modest case is that the disagreement is genuine. But leave that to the side. For it seems to me that Earth and Twin Earth need not diverge as they do in even this more modest example. To see this, it will be helpful to reflect on some of the literature produced by our most devout utilitarians—authors such as Peter Singer (1977) and Fred Feldman (1995). Singer rejects the Survival Lottery. He thinks that adopting it would fail to maximize utility because it would remove the incentive to take care of one's body. I do not think Singer is idiosyncratic in this regard. Utilitarians and Deontologists disagree about what we should do in uncommon, exotic, desert island type cases when we can save five people by killing one and no one outside the island will ever find out about it. Maybe they even disagree about whether in a case where an individual has the chance to perform an organ harvest he or she should do so. But they can agree that erecting an institution in which the practice of harvesting the organs of one to save the lives of many is standardized is wrong.

Consider Fred Feldman. On one of his formulations of utilitarianism, it is not just pleasure that must be maximized but desert adjusted pleasure. Harvesting the organs of a few to save the lives of many might, in the case of the Survival Lottery, maximize hedonic utility. But it would not maximize desert adjusted hedonic utility. And so this formulation of utilitarianism delivers the judgment that the Survival Lottery should not be adopted.

So Tc and Td disagree in this case about uncommon and exotic organ harvest cases and perhaps what an individual should do. (And perhaps Feldman's version of utilitarianism does not even disagree with Td about those cases.) However, Tc and Td agree in this case about what institutions and practices to promote. Neither theory advises us to adopt the Survival Lottery. So the inhabitants of future Earth and future Twin Earth can agree about rejecting the Survival Lottery. There need not be a divergence in moral institutions and practices here.

Similarly, I do not think the following piece of utilitarian reasoning about voting is idiosyncratic: Broyles, a utilitarian, is the President of Voting on Earth. He is in charge of all policies and institutions related to voting and is preparing for an upcoming election. Interestingly, Broyles does not believe that any average citizen is morally obligated to vote. Broyles reasons as follows: Consider any average citizen. He or she has two alternatives. One alternative is to vote. The other alternative is to not vote. It is probabilistically impossible that the outcome of the upcoming election will depend on just one vote. So either way, whether that individual citizen votes or not, the outcome of the election will not be affected. So there is no difference in utility between voting and not voting. So the potential voter is not morally obligated to vote.

Even though Broyles doesn't believe that average citizens are morally obligated to vote, he believes that, as President of Voting, he is morally obligated to promote institutions and practices that convince average citizens that they are morally obligated to vote. His reasoning is that if he fails to do so, significantly fewer average citizens will vote. And if that happens, that could affect the outcome of the election. Being a staunch believer in the utility democracy, Broyles believes that if more significant numbers of average citizens vote the more likely the outcome Thus, Broyles hires all the coolest hip hop artists and pop stars and launches a "Get out the Vote!" campaign.

Here we have another case where the institutions and practices of Earth and Twin Earth are very similar in spite of the fact that Tc-properties causally regulate 'right' on Earth and Td-properties causally regulate 'right' on Twin Earth. Tc and Td both recommend that authorities adopt policies that encourage individuals to vote. Td recommends this because each individual has a duty to vote. Tc recommends this because the promotion of such ideas maximizes utility even though no regular individual actually is obligated to vote. Tc and Td require different things from individuals when it comes to voting. But Tc and Td require the same institutions and practices be pursued by those in power.

3.8 Conclusion

Brink, Levy, and Merli have suggested that the disagreement intuition in Horgan and Timmons' Moral Twin Earth Argument will be undermined after reflection on the ways in which Earth and Twin Earth must diverge. I have argued that this is mistaken. The proponent of NMR must find some other way to respond to Horgan and Timmons' thought experiment.

CHAPTER 4

SYNTHETIC REDUCTIONISM AND HUEMERESQUE INTUITIONISM

Synthetic Reductionism holds that moral properties are identical to either natural kinds or theological properties. Huemeresque Intuitionism holds that moral properties are causally inefficacious and are fundamentally different from all other sorts of properties. In his interesting, important, and impressive book *Ethical Intuitionism*, Michael Huemer makes two critical points concerning Synthetic Reductionism. First, he attempts to rebut the common allegation that Synthetic Reductionism enjoys epistemological advantages that Humeresque Intuitionism lacks. Second, he claims that reflection on certain examples, together with a principle about reference, yields a powerful argument against Synthetic Reductionism. In this paper, I argue that Huemer is mistaken about both points. First, Huemeresque Intuitionism suffers from a serious epistemological problem—the Accidentalness Objection—that does not threaten Synthetic Reductionism. Second, the examples Huemer uses to motivate his argument against Synthetic Reductionism are misanalyzed and the principle about reference on which his argument rests is too rigid and excludes numerous cases of successful reference. I conclude by explaining how these considerations cast doubt upon the two central premises of *Ethical Intuitionism*'s overall argument.

4.1 Synthetic Reductionism

Huemer identifies his target as Synthetic Reductionism (SR). SR's proponents motivate SR on the basis of alleged analogies between natural kinds and natural kind terms, on the one hand, and moral properties and moral terms, on the other hand. Consider water, 'water', and *water*. Water is concrete stuff, 'water' is a natural kind term, and *water* is the natural kind *the property of being water*. Water fills lakes, oceans, cups, and bottles, 'water' is what English speakers use to refer to water, and *water* is the natural kind[40] that all and only concrete samples of water instantiate.

In recent years a certain story about water, 'water', and *water* has become fashionable. It is said that 'water' and 'H_2O' have distinct meanings, but that 'water' refers to H_2O and that the reference relation between 'water' and H_2O somehow entails that *water* and H_2O are identical properties.

Proponents of SR extend this fashionable story about natural kinds like *water* to moral properties like *goodness*, *badness*, *rightness*, and *wrongness*. They hold that 'good' refers to a natural kind or theological property and that *goodness* is identical to that natural kind or theological property. Furthermore, proponents of SR hold that *goodness* is causally efficacious. In addition, proponents of SR typically hold that the reference relation is to be analyzed in terms of causal relations between samples of moral properties and speakers.

As Huemer rightly points out, proponents of SR rarely offer concrete examples of these alleged identities between moral properties and natural kinds or theological properties. But there are exceptions and some proponents of SR do offer concrete theories. Richard Boyd, who holds that *goodness* is a consequentialist homeostaitic property cluster[41], is a naturalist example of such a proponent. Robert Adams, who holds that *goodness* is the *property of being similar to God in a certain way*[42], is a theological example of such a proponent.

4.2 Huemeresque Intuitionism

[40] Here I am attributing to proponents of SR the claim that natural kinds are properties (of a special sort).

[41] What is a consequentialist homeostatic property cluster? See Boyd (1988) and Rubin (200x).

[42] What way of being similar to God is that? See Adams (1999) and Hill (forthcoming).

Huemeresque Intuitionism (HI) rejects the surprising property identities put forward by proponents of SR. Moral properties are not identical to any natural kinds or theological properties. They are irreducible, fundamentally different from all other sorts of properties, and causally inefficacious.

4.2.1 Huemeresque Intuitions: What They Are and Their Role in Justification

HI includes an admirably clear and precise account of intuitions and their role in justification. Intuitions are a certain sort of appearance (Huemer 2005, pp. 99-102; Huemer 2007, pp. 30-1). When one utters a sentence like ‘It seems to me that *p*’ or ‘It appears that *p*’ or ‘It is obvious that *p*’ one has reported an alleged appearance. Appearances have propositional content. They can be perceptual, mnemonic, introspective, or intellectual. Appearances are distinct from beliefs and dispositions to form beliefs. However, they sometimes lead to the formation of beliefs and all judgments that *p* are based in some way on appearances that *p*. Appearances sometimes conflict with one another. Appearances vary in strength.

If an appearance is intellectual rather than perceptual or mnemonic or introspective, then that appearance is an intuition (Huemer 2005, pp. 100-2). Examples of intuitions are the intellectual appearance that time is one-dimensional and ordered, the intellectual appearance that nothing can be both completely red and completely blue at the same time, the intellectual appearance that 2 + 2 = 4 and the intellectual appearance that a particular argument form is valid. An ethical intuition is an intuition that is such that its propositional content is ethical. An example is the intuition that torturing babies just for fun is wrong.

Huemer (2005, p. 99; 2007, p. 30) combines the above account of intuitions and appearances with the Principle of Phenomenal Conservativism (PC):

PC: If it appears to *S* that *p*, then, in the absence of defeaters, *S* thereby has at least some degree of justification for believing that *p*.

Since ethical intuitions are special sorts of appearances, PC delivers the judgment that they, in the absence of defeaters, confer some degree of justification for believing their propositional content. For example, most people have the intuition that torturing babies just for fun is wrong. So PC judges that, in the absence of defeaters, most people have some degree of justification for believing that torturing babies just for fun is wrong. HI, then, holds that ethical intuitions, combined with the absence of genuine defeaters, and the truth of the propositional content of those intuitions, together with whatever needs to be added to true belief and justification to produce knowledge, provides one with knowledge of the sorts of moral properties discussed above. This is *a priori* knowledge. It is obtained in the same way that mathematical and logical knowledge is.

4.2.2 Identifying Targets

Ethical Intuitionism is a rich and complex book. It contains numerous metaethical theses and combines them in interesting ways. It seems to me, however, that these theses are not entirely interdependent and it is not essential that they be put together in the exact way that Huemer has done. For this reason it will be prudent to discuss some of the most important claims made by Huemer and to identify the claims I hope to target together with the claims for which I have no criticisms.

Intuition: Huemer claims that moral theorizing can proceed successfully only if moral theorists rely on moral intuitions. I have no criticisms of this claim. I believe it to be true.

Phenomenal Conservativism: Huemer advocates PC. He claims that, in the absence of defeaters, if it appears to S that P, then S has some degree of justification for believing that P. While I stop short of giving PC my full and whole hearted endorsement, I find Huemer's defense of PC to be compelling and plausible. I have no criticisms to offer here.

Intuition Implies Epistemic Equivalence: Huemer claims that given that the use of moral intuitions is required for successful moral theorizing, SR enjoys no epistemological advantages over HI. I believe he is mistaken. This claim is a target of my criticisms.

Intuition and Reference Imply SR is False: Huemer claims that certain intuitions, combined with reflection on certain examples and a principle about reference, show that SR is false. I am not convinced. This claim is a target of my criticisms.

4.3 SR versus HI with respect to Moral Knowedge

Some authors claim that SR is better able to account for moral knowledge than forms of Intuitionism such as HI. A common motivation for this claim, and the motivation that Huemer targets, is that intuitions are "spooky" and "mysterious" and somehow epistemologically suspicious. So a theory that does without appeals to intuitions is to be preferred over a theory that requires such appeals. While HI holds that moral claims are justified by way of intuitions, SR holds that moral claims are justified only on the basis of empirical and scientific inquiry. So HI requires an appeal to intuitions but SR does not. Thus, SR enjoys a considerable epistemological advantage over HI.

Huemer attempts to show that this alleged advantage of SR is illusory. In short, the argument is this: Proponents of SR tell a number of different stories about how moral claims are justified solely on the basis of empirical and scientific inquiry. Huemeresque scrutiny, however, reveals that each of those stories, prominent though they may be, either relies on some confusion or makes an implicit appeal to intuition. Proponents of SR are free to help themselves to appeals to intuition. But then they cannot claim that SR enjoys any epistemological advantage over HI. Both theories appeal to intuition. So they are epistemologically equivalent. Huemer makes the point this way:

> [T]he synthetic reductionists' account of moral knowledge fails: no theory about the nature of goodness can be known in any manner analogous to how the above theories of heat, water, and sound are known. In the end, the synthetic reductionist will have to appeal to ethical intuition, just as the intuitionists do. This deprives their position of one of its central alleged advantages.

In more detail, Huemer's argument runs as follows: Proponents of SR can accommodate moral knowledge only if the following claim is true:

INT: Moral theorists have intuitions concerning which particular things and kinds of things instantiate moral properties and those intuitions provide moral theorists with justification for believing the propositional content of those intuitions.

However, if INT is true, then SR is on the same epistemological footing as Intuitionist theories such as HI. So SR does not enjoy any advantage over Intuitionism. Or, put another way:

(1) SR accommodates knowledge that *goodness* is identical to *G* only if it is combined with INT.
(2) If SR accommodates knowledge that *goodness* is identical to *G* only if it is combined with INT, then SR enjoys no epistemological advantages over HI.
(3) Therefore, SR enjoys no epistemological advantages over HI.

Concerning premise (1), remember that SR is supposed to be an extension of the sort of reasoning that motivates the adoption of natural kind property identities like *water* is *H2O*. According to Huemer, an essential component of this reasoning is that, prior to theorizing about *water*, people know which samples of stuff instantiate *water* and which samples of stuff do not. Huemer makes the point this way:

> The reason why we believe, for example, that water = H2O is, roughly, that [among other things] we have independent (that is, pre-theoretical), direct awareness of the presence of water…. [Likewise] we were able to discover that heat = molecular kinetic energy, only because we could first identify which things were hot.

The relevant issue, then, is whether, prior to theorizing about moral properties like *goodness*, people know which samples of stuff instantiate *goodness* and which samples of stuff do not. In the case of *water,* pretheoretical knowledge of which things instantiate *water* and which do not is obtained by way of observation. To observe a property, instances of that property must effect one's senses. But moral properties, Huemer says[43], "do not look like anything, sound like anything, feel (to the touch) like anything, smell like anything, or taste like anything." Huemer considers four attempts to account for the relevant moral knowledge by prominent adherents of SR. He considers the idea that the relevant moral knowledge is in fact obtainable by way of direct observation, inference to the best explanation from other things that are observable, the testability of moral claims, and the unifying power of moral explanation. Huemer argues convincingly that each of these strategies, prominent as they are, cannot offer a plausible account of the sort of moral knowledge in question. The accounts offered by proponents of SR concerning how we know which samples of stuff instantiate *goodness* and which do not are all implausible. And without such knowledge, the proponent of SR cannot make a case for the alleged property identities SR holds to be true. So even if, as SR holds, *goodness* were identical to a natural kind or theological property, we could never know it. Huemer makes the point eloquently:

> If we are to take the synthetic reductionist's analogies seriously, then, we should say that we have independent, direct awareness of the presence of goodness…. If we have no pre-theoretical knowledge of which things are good, then we cannot discover that Good = *N* in any analogous way. The situation would be like the following scenario: someone tells you that there is an unobservable property called 'torfness'. You are given no initial information about the nature of

[43] 85

> torfness or which things in the world, if any, are torf. You are told only that torfness is reducible to some scientific property. Your assignment: figure out which scientific property torfness is reducible to. It seems clear that you do not have enough information to carry out the assignment. Likewise, with no initial information about the nature of goodness or which things have it, we would have no basis for framing any hypotheses as to what natural property goodness might be reducible to…. One might be tempted to appeal to common sense beliefs about which things are good—but that would be to revert to a mistake we have earlier criticized. Since at this point the reductionist has not yet successfully explained how people can access moral reality, he is not entitled to assume that our common sense moral beliefs are reliable.

Now, since their own accounts of the relevant moral knowledge do not work, proponents of SR might claim that such knowledge is obtained by way of intuitions. Prior to theorizing, people have intuitions that particular objects or states of affairs instantiate *goodness*. Combined with some intuitionist theory of justification such as Huemer's PC, and in the absence of defeaters, the relevant beliefs are justified. Given that the degree of justification is strong enough, that they are true, and barring Gettier cases, people know the propositional content of these intuitions. In other words, the proponent of SR can account for moral knowledge by adopting INT.

To sum up, the defense of premise (1) is this: If SR is true, then knowledge that a moral property like goodness is identical to a natural or theological property, *G*, is obtained in the same way that knowledge about natural kind property identities, such as *water* is identical to H_2O, are allegedly obtained. But knowledge that *water* is H_2O is obtained only if moral theorists know, prior to theorizing, which things instantiate *water* and which things do not. Similarly, knowledge that a moral property, like *goodness*, is identical to some natural kind or theological property, *G*, is obtained only if moral theorists know, prior to theorizing, which things instantiate *goodness* and which do not. But, as Huemer convincingly argues, the accounts of such pretheoretical knowledge offered by proponents of SR are all implausible. And any alternatives they have put forward that do not rely on such pretheoretical knowledge are all equally implausible.

The only option left is INT. So SR can accommodate knowledge that *goodness* is identical to *G* only if SR is combined with INT.

The motivation for premise (2) is seemingly straightforward. SR is often alleged to better accommodate moral knowledge than intuitionist views like HI. But if SR is combined with INT, then SR's epistemological fate rests with HI's. Both rely on INT. If INT is an unreasonable and "spooky" epistemological thesis, then both HI and SR are in trouble. If INT is reasonable and can be sufficiently "unspookified" for the ontologically uptight naturalistic versions of SR, then the proponent of HI should be able to help herself to INT as well. So if SR can accommodate knowledge that *goodness* is identical to *G* only if it is combined with INT, then SR enjoys no epistemological advantage over Intuitionist views like HI.

4.4 A Reformulation of Huemer's Argument

It seems to me that there is a much more interesting way to formulate Huemer's argument. Consider premise (1). The defense of this premise starts with the assumption that SR is not combined with INT and then, given that assumption, derives the following claim:

(A) SR cannot accommodate knowledge of which things instantiate *goodness* and which do not.

(A) is then used to establish the claim below:

(B) SR cannot accommodate knowledge that *goodness* is identical to *G*.

and then (B), in light of the assumption made at the start of the defense, is used to establish premise (1). What I find odd about this is that it seems to me that (A) is much more threatening to SR than (B).

To see this, contrast two theories, T_1 and T_2. Never mind all the precise details of these theories. Just note that, first, T_1 and T_2 are both theories of *goodness* and they are both forms of SR. So they are both committed to the claim that *goodness* is identical to some natural kind or some theological property. Second, note that T_1 and T_2 have the following consequences: T_1 has the consequence that normal people and moral theorists alike know which things instantiate *goodness* and which do not. But no normal person or moral theorist could know which natural kind or theological property *goodness* is identical to. T_2, on the other hand, has the consequence that neither normal people nor moral theorists know which things instantiate *goodness* and which do not. But they do know which natural kind or theological property *goodness* is identical to. If, going only on the information above, I had to choose between T_1 and T_2, I would, without hesitation, pick T_2. I think it is a huge defect in a theory of *goodness* if it has the result that no one knows which things instantiate *goodness* and which do not. Leaving to the side nihilism and skepticism, it is obvious that people know which things instantiate *goodness* and which do not. An allegedly non-nihilistic and non-skeptical theory that cannot accommodate this obvious fact has a serious problem. On the other hand, supposing SR is true[44], I think it is a much lesser defect in a theory of *goodness* if it has the result that no one could ever know which natural kind or theological property is identical to goodness. This is mainly because it seems clear to me that even if SR is true, no one yet knows (and no one may ever know) which natural kind or theological property *goodness* is identical to. Proponents of SR have made very few concrete proposals about which

[44] Note that if SR isn't true and, say, HI is, then the fact that one could never know which natural kind or theological property is identical to goodness isn't a problem at all. After all, it isn't identical to any such property!

natural kind or theological properties *goodness* is identical to.   It is all just speculative theorizing.

In sum, my point is this: It is a big defect in a theory if, like T_2, it has the consequence that we don't know something that we obviously do know like which things instantiate *goodness* and which do not. It is a much smaller defect in a theory if it, like T_1, has the consequence that we couldn't know something that we don't know and may never know like which natural kind or theological property *goodness* is identical to. T_2 accommodates actual knowledge of commonsense moral claims but not possible knowledge of speculative philosophical claims. T_1 accommodates possible knowledge of speculative philosophical claims but not actual knowledge of commonsense moral claims. Just going on this information, T_2 is the better theory.

In light of all of this, it seems to me that a more interesting formulation of Huemer's argument would be this:

(1) SR accommodates knowledge about which things instantiate *goodness* and which do not only if it is combined with INT.
(2) If SR accommodates knowledge about which things instantiate *goodness* and which do not only if it is combined with INT, then SR enjoys no epistemological advantages over HI.
(3) Therefore, SR enjoys no epistemological advantages over HI.

For this reason, in my reply to Huemer, I will target this formulation of the argument rather than the one explicitly presented in the text. However, even if you disagree with me and are more worried about the original argument, what I say about the reformulated argument will, pretty straightforwardly I think, apply to the original formulation as well.

4.5 Problems for Huemer's Argument

As I see it, Huemer's argument suffers from one main problem. Premise (2) is false. Huemer argues convincingly that SR can accommodate the relevant moral knowledge only if it is combined with INT. But it does not follow that SR is not epistemologically superior to HI. Some SR and INT combinations avoid a defeater that HI does not avoid. Such forms of SR enjoy an epistemological advantage over HI.

4.5.1 Premise (2) of Huemer's Argument is False

In this section I will show that premise (2) of Huemer's argument is false. Suppose that one grants Huemer the antecedent of (2). So SR accommodates knowledge about which things instantiate *goodness* and which do not only if it is combined with INT. The consequent of (2) does not follow. To see this, suppose we combine SR with INT. In fact let us not only combine SR with INT. Let us combine SR with Huemer's PC. What I will show below is that Huemer's form of intuitionism suffers from a defeater that certain forms of SR combined with INT and PC do not. So these formulations of SR enjoy an epistemological advantage that HI lacks. Thus, premise (2) of Huemer's argument is false.

4.5.2 The Accidentalness Objection

As we have seen, one motivation for the claim that SR enjoys epistemological advantages over HI is that SR need not appeal to "spooky" intuitions. I agree with Huemer that this motivation is inadequate. For one thing, I am a big fan of spookiness. Contrary to popular misconceptions, a theory's being spooky in some respects is no strike against it. For another thing, I find Huemer's critique of SR's alternatives to INT to be decisive. I just don't understand how we can know when we have hit our theoretical targets unless we assume, with INT, that our intuitions about which things instantiate moral properties and which things do not are reliable. Huemer's torfness example illustrates the point nicely. Even so, there are other ways of motivating the claim that SR enjoys an epistemological advantage over HI. Perhaps the most powerful motivation is what I will call the "Accidentalness Objection." Huemer (2005) articulates the Accidentalness Objection this way:

> There is a… general condition on knowledge that everyone in epistemology accepts: I know *p* only if it is not a mere accident (not a matter of chance) that I am right about whether *p*. The challenge for the moral realist, then, is to explain how it would be anything more than chance if my moral beliefs were true, given that I do not interact with moral properties.

We can state Huemer's concern as follows: If HI is true, then moral properties are causally inefficacious. If moral properties are causally inefficacious, then any true belief about the truth-value of a moral proposition anyone has is accidentally true. However, if a subject's belief that p is accidentally true, then that subject does not know that p. So if HI is true, then no one knows the truth-value of any moral proposition. But some people do know the truth-values of some moral propositions. So HI is false.

Accidentally Adequately Grasping

It is sometimes claimed that Moral Non-Naturalism (MNN) cannot accommodate moral knowledge. Even if our moral beliefs were true, it is alleged, they would be true in merely accidental way. The heart of Michael Huemer's interesting, important, and impressive book *Ethical Intuitionism* defends MNN from this allegation. Huemer sketches a theory of *a priori* knowledge and claims that if a subject forms beliefs about moral properties in accordance with his sketch, then that subject's beliefs about that property are not problematically accidentally true. In this paper, I describe a simple case in which a subject forms beliefs about *goodness* in accordance with Huemer's sketch but that subject's beliefs about *goodness* are true in a merely accidental way. I then try to fill out Huemer's sketch in a way that might enable him to accommodate this case. I argue that the filled out sketch is also unsuccessful. It faces familiar problems Huemer himself raises against other metaethcial theories and eventually leads back to the problem of accidentalness. I conclude that the theory of *a priori* knowledge sketched by Huemer is of no help to the proponent of MNN.

4.5.3 Adequately Grasping

Huemer's sketch includes an account of adequate grasping[45].

A subject adequately grasps a universal if and only if that subject's grasp of that universal is consistent, clear, and determinate.

For a subject's grasp of a universal to be consistent is for that universal to be possibly instantiated. For a subject's grasp of a universal to be clear is for that subject to be able to successfully distinguish between that universal and other similar universals. For a subject's grasp of a universal to be determinate is for there to be few or no borderline cases in which the subject is unable to successfully determine whether or not that universal is instantiated.

So what work is the account of adequate grasping supposed to do? Huemer's (pp. 123-7) answer is that:

> There is a… condition on knowledge that everyone in epistemology accepts: I know *p* only if it is not a mere accident (not a matter of chance) that I am right about whether *p*. The challenge for the [proponent of MNN], then, is to explain how it would be anything more than chance if my moral beliefs were true…. [A]dequately grasping a universal cannot cause false intuitions about it…. [W]hen one's intuitions are caused (only) by clear, consistent, and determinate understanding—

[45] Humer states (p. 125) this account in the following passage:
In having concepts, we grasp (understand) universals. To have the concept of yellow is to understand (at least partly) what yellow is. Grasping comes in degrees: one may grasp something better or worse. An *adequate* grasp of a universal is a concept that is:
i) *Consistent*. For instance, the concept of a round square is inconsistent; accordingly, it does not count as an adequate grasp of a universal. Subtler inconsistencies exist, such as that perhaps involved in the concept of the largest prime number.
ii) *Clear*, as opposed to confused. For instance, someone who has read a little about chaos theory may think that 'chaos' is like 'randomness'. He may, that is, fail to distinguish these two concepts. In that case, his concept of chaos is confused and is not an adequate grasp of the nature of chaos.
iii) *Determinate*, as opposed to vague or unsettled. Imagine someone arguing about abortion. You tell him about the RU-486 pill and ask whether it counts as abortion. He doesn't know. In this case, his concept of abortion is indeterminate; the criteria for applying the concept are unsettled in his mind.
The above characteristics come in degrees; we say a person's understanding is adequate if his concept has these traits to a high degree.

> the internal process by which one forms beliefs guarantees their truth…. The above… addresse[s]… the problem of why, even if true and justified, our moral beliefs would not be merely *accidentally* true.

So the idea is that adequately grasping a universal ensures that one's beliefs about that universal are true. And if adequate grasping ensures truth in this way, then moral beliefs resulting from an adequate grasp of *goodness* or *badness* or *rightness* or *wrongness* are not accidentally true.

4.5.4 Accidentally Adequately Grasping

Huemer's account of adequate grasping does not do the work it is supposed to do. While adequate grasping ensures truth, it does not ensure non-accidentalness. To see why, consider the following example: ROBOTRON 5000 has always had strong moral convictions. From the time he left the factory where he was built, for instance, he has had strong opinions about *goodness*. He enjoys making lists of properties and distinguishing them from *goodness*. He has a very precise view about when *goodness* is instantiated and when it is not. For any case whatsoever, he will tell you with complete confidence that *goodness* is definitely instantiated or that it is definitely not instantiated in that case. There are never any puzzling or borderline cases for him. He is always confident that *goodness* is instantiated or fails to be instantiated. For many years now, he has wondered how he got these strong moral convictions.

One evening ROBOTRON 5000 and his designer walk into a bar. Given his strong convictions and given his interest in the origins of his convictions, ROBOTRON 5000 decides to take this opportunity to ask his designer how he got them. The designer tells him that around the same time he built ROBTRON 5000, he was tinkering with a state of the art random truth-value assigner (RTA). He input all propositions asserting that *goodness* is distinct from or identical to various properties, propositions such as

‘*Goodness* is identical to *electronic-musician-ness*’, into the RTA. He input all propositions asserting that *goodness* is (or is not) instantiated in various cases, propositions such as ‘*Goodness* is not instantiated by the Holocaust’, into the RTA. Then the RTA went to work and randomly assigned truth-values to each of those propositions. The designer then programmed that assignment of truth-values into ROBOTRON 5000’s belief system. That is where ROBOTRON 5000 got his strong moral convictions.

Now, as a matter of sheer luck, the RTA’s assignment of truth-values happened to match the actual truth-values of all the relevant propositions about whether *goodness* is identical to some property or other and whether *goodness* is instantiated in some case or other. A consequence of this is that ROBOTRON 5000’s grasp of *goodness* is clear. He can distinguish *goodness* from other properties. If you ask him “Is *goodness* identical to *electronic-musician-ness*?” he will correctly answer “No.” For any property you give him, he will give you the right answer about whether that property is identical to *goodness*. Another consequence is that ROBOTRON 5000’s grasp of *goodness* is determinate. There are no borderline cases whatsoever in which ROBOTRON 5000 cannot tell whether *goodness* is instantiated or not. For any case you give him, he has got a strong opinion about whether *goodness* is instantiated. And he always gets it right. Finally, ROBOTRON 5000 has a consistent grasp of *goodness*. *Goodness* isn’t like *round-square-ness*. It is possibly instantiated. And none of ROBOTRON 5000’s beliefs about *goodness* are inconsistent. Thus, ROBOTRON 5000 has an adequate grasp of *goodness*. And his adequate grasp guarantees that his beliefs about *goodness* and all other moral properties are true.

Poor ROBOTRON 5000. Before he walked into the bar that night he might have been justified in maintaining his strong moral convictions. Now, however, he knows that his moral beliefs are at best only accidentally true. He has a defeater for those beliefs. He adequately grasps *goodness* and all other moral properties. That adequate grasp certainly guarantees the truth of his moral beliefs. But it is still an accident. ROBOTRON 5000's moral beliefs are now unjustified. And they never counted as knowledge.

Given MNN, ROBOTRON 5000's story does not seem very different from our own. On MNN, the evolutionary processes that formed our moral convictions are just as disconnected from *goodness* as ROBOTRON 5000's RTA is. Reflection on this makes vivid the worry that if MNN is true, then our moral beliefs are now unjustified. And even if they were once justified, and even if we got lucky and they are true, we never knew that they were true. Huemer's discussion of adequate grasping does not lessen the force of this worry.

4.5.5 The Response from Plenitude

One might think I have ignored an important feature of Huemer's theory. Huemer (p. 126) maintains that universals are plentiful and that they are necessary:

> I propose that having a clear, consistent, and determinate concept is *sufficient* for one's grasping a universal or universals. There is no possibility of one's failing to refer to anything (universals are plentiful in this sense, and their existence is necessary).

Once one fully appreciates this feature of Huemer's theory, one might think, the intuition that ROBOTRON 5000's moral beliefs are formed in a problematically accidental way will be undermined. This response is suggestive. It has a ring of truth to it. But it stands in need of development. In what follows, I will try to offer such development on Huemer's behalf.

Adequate Grasps and Adequate Concepts: As we have seen, Huemer applies the terms 'clear', 'consistent', and 'determinate' to a subject's grasp of a universal. In the passage just quoted, however, Huemer applies these terms to a subject's concept. Now, one might think that these are just two ways of talking about the same thing. But two features of the passage in question strongly suggest otherwise. First, Huemer says that having an adequate concept is sufficient for grasping a universal. This would be trivial if having an adequate concept and having an adequate grasp were identical. Second, given the definition of adequate grasping, there will be a universal that corresponds to a subject's grasp regardless of how plentiful universals are. But, in the relevant passage, considerations of plenitude are supposed to explain why having an adequate concept guarantees that that concept corresponds to a universal. Again, this would be a trivial point if having an adequate concept and having an adequate grasp were identical. So, assuming that Huemer is not here concerned to make trivial points, having an adequate concept and having an adequate grasp must be distinct.

Huemer gives a precise account of adequately grasping. But he says nothing about what it is to have an adequate concept. Perhaps what Huemer has in mind is this: Having an adequate concept is just like having an adequate grasp with the exception that successful correspondence is built into the definition of having an adequate grasp but is not built into the definition of having an adequate concept. Here is what I mean:

A subject has an adequate concept of a candidate universal if and only if that subject's concept is consistent, clear, and determinate.

For a subject's concept to be consistent is for all of that subject's beliefs about that candidate universal to be consistent. For a subject's concept to be clear is for that subject to judge that that candidate universal is distinct from other similar candidate universals.

For a subject's concept to be determinate is for there to be few or no cases in which that subject is unable to form a judgment about whether that candidate universal is instantiated[46].

Now, notice the differences between having an adequate grasp and having an adequate concept. We switched from talk of universals to talk of candidate universals. We switched from talk of a universal being possibly instantiated to talk of a subject's beliefs about a candidate universal being consistent. We switched from talk of successfully distinguishing between universals and of successfully judging that a universal is instantiated to talk of forming beliefs about whether a candidate universal is distinct from other candidate universals and forming beliefs about whether a candidate universal is instantiated. In the case of adequate grasping, successful correspondence is built into its definition. In the case of having an adequate concept successful correspondence is not.

This accommodates the interpretive constraints identified in the relevant passage. First, it accommodates the constraint that having an adequate grasp and having an adequate concept are distinct. Second, it accommodates the constraint that it is not trivial considerations, but rather considerations of plenitude, that guarantee that having a universal will always correspond to an adequate concept.

[46] Huemer (p. 126) says that an adequate concept (grasp) can be accompanied by other things that mitigate the epistemic benefits of adequacy:

Notice that the claim is not that all intuitions are true. Nor is the claim that all intuitions of a person who adequately grasps the relevant concepts are true. As we shall see… there are many ways we can go wrong. The claim is that *adequately grasping a concept* cannot itself cause a mistake. In other words, if a person has a false intuition, this false intuition must be caused by something else—by misunderstanding, bias, confusion, or the like. It is consistent with this that there be many false intuitions.

Let us say that a subject's concept (grasp) is *purely* adequate when that subject has an adequate concept (grasp) and none of that subject's clarity or determinacy judgments are caused by "misunderstanding, bias, confusion, and the like." Let us further say that purity comes in degrees. The less misunderstanding, bias, confusion, and the like, the purer the adequate grasp. I stipulate that all the robots discussed in this paper have purely adequate concepts and grasps.

Clarity and Determinacy Judgments: Huemer analyzes adequacy in terms of sentences of the form ‘U is (or is not) distinct from U*’ where U and U* are (candidate) universals. He also analyzes adequacy in terms of sentences of the form ‘U is (or is not) instantiated in C’ where U is a (candidate) universal and C is a case. In what follows, it will be useful to introduce names for such sentences. Call a sentence with the first form a ‘clarity judgment’ and a sentence with the second form a ‘determinacy judgment’.

Being True Of: Huemer talks about universals corresponding to concepts. On Huemer’s behalf, I suggest analyzing correspondence in terms of a *being true of* relation between concepts of candidate universals and universals. Consider a clarity or determinacy judgment that a subject associates with a concept of a candidate universal U. Remove the name of U from that judgment and replace it with the name of a universal U*. If the proposition expressed by the modified sentence is true, then say that the judgment that subject associates with her concept of U is true of U*.

For example, consider ROBOTRON 5000’s clarity judgment that ‘*Goodness* is distinct from *pleasurable-ness*’. Now, consider a universal, say *watery-ness*. The clarity judgment ROBOTRON 5000 associates with his concept of *goodness* is true of *watery-ness*. For ROBOTRON 5000 judges that *goodness* is distinct from *pleasurable-ness*. And it is true of *watery-ness* that it is distinct from *pleasurable-ness*. So the relevant clarity judgment ROBOTRON 5000 associates with his concept of *goodness* is true of *watery-ness*.

Correspondence and Plenitude: Now we are in a position to propose accounts of correspondence and plenitude on Huemer's behalf[47]:

CORRESPONDENCE: A universal corresponds to a subject's concept $=_{def}$ all clarity and determinacy judgments that subject associates with that concept are true of that universal.

PLENITUDE: For any adequate concept, there is a universal that corresponds to that concept.

The Response from Plenitude: With the doctrine of PLENITUTDE now in hand[48], we are in a position to offer a developed version of the response to my objection. Consider again ROBOTRON 5000. My objection, one might argue, only has intuitive force if we assume that ROBOTRON 5000 could have had an adequate concept of a candidate universal without there being any universal matching that concept. In such a case it would be an accident that his concept ends up matching a universal. That would be problematic. But, given PLENITUDE, one cannot have an adequate concept of a candidate universal without the existence of a universal that corresponds to that concept. So, as long as the RTA provides ROBOTRON 5000 with an adequate concept, then whatever that concept turns out to be, there will be some universal or other that matches his concept. Thus, once PLENITUDE is taken into account, the intuition that ROBOTRON 5000's beliefs about *goodness* were formed in a problematically accidental

[47] In the passage in question, Huemer also seems to be suggesting a principle about reference. He says "There is no possibility of one's failing to refer to anything. Universals are plentiful in this sense." I take him to be suggesting this:
REFERENCE: For any term that a subject associates with a concept, if that subject's concept corresponds to a universal, then that subject's use of that term refers to that universal.

[48] I take PLENITUDE to be suggested by the following passage: "So the intrinsic characteristics of a concept sometimes are sufficient for its constituting an adequate understanding of the nature of a universal. Furthermore, adequately grasping a universal cannot cause false intuitions about it. Therefore, in some cases—namely, when one's intuitions are caused (only) by clear, consistent, and determinate understanding—the internal process by which one forms beliefs guarantees their truth. Some will say that 'guarantees' is too strong; all I need say is 'renders highly probable'. But I think the 'guarantee' claim is correct."

way disappears. It is an accident that he has the concept of *goodness*. But given that he has it, and given PLENITUDE, it is no accident that there is a universal corresponding to it. Thus, the sort of accidentalness present in this case is not problematic.

To illustrate, consider another robot—UNTERTRON. She received an adequate concept of the candidate universal *being-dry-land-below-sea-level-ness* through the deliverances of an RTA. She associates the term 'unterland-ness' with this candidate universal. In fact, there is such a universal. And so she adequately grasps *unterland-ness*. Like ROBOTRON 5000 does with *goodness*, UNTERTRON goes around successfully identifying portions of land that instantiate *unterland-ness* and other things that do not instantiate *unterland-ness*.   In addition, she makes lists of similar universals and successfully distinguishes them from *unterland-ness*. Now, given the way the RTA works, it is accidental that UNTERTRON adequately grasps *unterland-ness*. She could have easily ended up with an adequate concept of some other candidate universal such as *being-dry-land-at-least-one-foot-below-sea-level-ness*. If she had, she would have associated the term 'unterland-ness' with that different candidate universal and her classifying activities would have differed accordingly. So it is an accident which candidate universal UNTERTRON has an adequate concept of or that she has such a concept at all. But given that she does end up with such a concept, and given PLENITUDE, her adequate concept is guaranteed to be an adequate grasp and there is going to be some universal or other that matches whatever concept the RTA gave her. So, although her unterland beliefs were formed by accident, they were not formed in a problematically accidental way. Thus, UNTERTRON's unterland knowledge is not undermined. In the same way, ROBOTRON 5000's moral knowledge is not undermined[49].

4.5.6 Disagreement, Error, and the Return of Accidentalness

The problem with this response is that it renders Huemer's theory vulnerable to objections he repeatedly (pp. 50, 52-3, 56, 63-4) wields against other metaethical theories. For example, Huemer's theory is unable to accommodate moral disagreement. Consider two robots—UTILOTRON and KANTOTRON. Both robots received, by way of an RTA, adequate concepts with which they associate the term 'wrongness'. But they received distinct concepts. UTILOTRON received a utilitarian concept of wrongness. All of his clarity and determinacy judgments match the deliverances of utilitarianism. KANTOTRON, on the other hand, received a Kantian concept of wrongness. All of his clarity and determinacy judgments match the deliverances of Kantianism. Furthermore, the utilitarian and Kantian concepts of wrongness are both consistent.

Now, suppose that UTILOTRON and KANTOTRON witness an organ harvest. The organ harvest maximized utility by killing one to save the lives of two. So utilitarianism judges that it was not wrong. However, the organ harvest required killing an innocent and unwilling donor. This involves treating a person merely as a means. So Kantianism judges the action to be wrong. In response, KANTOTRON says 'The organ harvest is wrong' and UTILOTRON says 'It is not the case that the organ harvest is wrong'. The two robots believe that they are disagreeing with each other. Each robot tries to persuade the other. Neither robot changes his mind.

[49] I am assuming here, on Huemer's behalf, that introspection enables a subject to reliably judge whether his or her concepts are adequate. Huemer seems to endorse this assumption. "[I]ntrospection" he says "reveals that we sometimes understand concepts" (pp. 125-6).

Now, given PLENITUDE, there is a universal that corresponds to each of UTILOTRON and KANTOTRON's concepts. There is a universal that matches the utilitarian concept of wrongness and a universal that matches the Kantian concept of wrongness. Both universals exist. UTILOTRON uses 'wrongness' to refer to the former and KANTOTRON uses 'wrongness' to refer to the latter. Thus, UTILOTRON and KANTOTRON are not really disagreeing. When UTILOTRON says 'It is not the case that the organ harvest is wrong' what he says is true. When KANTOTRON says 'The organ harvest is wrong' what he says is true. They are not contradicting each other. They are just talking past each other. Their apparent disagreement is not genuine. But this is absurd. UTILOTRON and KANTOTRON are obviously disagreeing. They are both talking about *wrongness*. They just have different theories about what *wrongness* is. Thus, Huemer's theory, on the present interpretation, is false. It cannot accommodate moral disagreement.

Consider another robot—NIETZSHOTRON. He received an adequate concept from an RTA that he associates with the term 'goodness'. That concept is the Nietzschean "Will to Power" concept of *goodness*. Accordingly, he accepts the determinacy judgment that 'Domination and control of others instantiates *goodness*.' Given PLENITUDE, there is a universal that matches NIETZSHOTRON's concept. Thus, his determinacy judgment about the *goodness* of dominating and controlling others is true. But this is absurd. NIETZSHOTRON's judgment is false. NIETZSHOTRON's concept of *goodness* is a mere distortion and perversion of the *goodness* we all know and love. He has a mistaken theory of *goodness*. Thus, Huemer's theory, on the present interpretation, is false. It cannot accommodate moral error. Again, these are familiar objections that Huemer repeatedly employs against other metaethical theories in *Ethical Intuitionism*.

Let me take stock. I think these considerations about disagreement and error point back to the problem of accidentalness. Moral properties are special. They “carve at the joints” in a way that non-moral properties with partially overlapping extensions do not. If there are universals corresponding to both Kantian and utilitarian concepts of wrongness, at most only one of them carves at the joints. If there is a universal corresponding to the Nietzschean concept of *goodness*, it does not carve at the joints in the way that *goodness* does. In the case of UNTERTRON, neither unterland-ness nor any universal with a similar extension she might have carves at the joints. So there is no worry that she accidentally got the one concept in the neighborhood that matches a universal that carves at the joints but could have easily gotten one that does not. ROBOTRON 5000 is in a very different position, he just happened to grasp *goodness*. *Goodness* carves at the joints. Other universals with partially overlapping extensions do not. If we adopt PLENITUDE, then the worry is not that it is accidental that there is a universal that matches ROBOTRON 5000’s concept at all. The worry is instead that it is only an accident that his concept matches the one universal in the neighborhood that carves at the joints. That is problematic. I want to know why we are, at best, not like ROBOTROBN 5000. If PLENITUDE is true, I want to know how it could be anything other than an accident that our concept of *goodness* corresponds to a universal that carves at the joints. The doctrine of PLENITUDE does not help me answer this question. We are back to the problem of accidentalness.

4.6 The Accidentalness Objection is no Problem for SR

Huemer claims to have shown that SR does not enjoy any epistemological advantages over HI. We are now in a position to see that he is mistaken. First, consider Robert Adams' version of the Divine Command Theory. Huemer classifies Adams' version of the Divine Command Theory differently than I do. He classifies it as a form of subjectivism. I classify it as a form of SR. There is truth in both classifications. But how one decides to classify Adams' theory is irrelevant to the present discussion in light of the following two points: First, Huemer explicitly targets Adams' version of the Divine Command Theory in *Ethical Intuitionism* and, second, Huemer claims that the Divine Command Theory does not enjoy any epistemological advantage over HI for the same reason he claims SR generally enjoys no such advantage. He says, for example, this:

> The Divine Command theorist would have to posit some sort of access to moral facts independent of our knowledge of God's will, which would seem to surrender one of the major advantages of a Divine Command theory—namely, its ability to explain moral knowledge. For example, the Divine Command theorist might be driven to posit a faculty of moral intuition—but then it is unclear how his theory would be better than traditional ethical intuitionism.

In light of the discussion of ROBOTRON 5000 above, it is clear that Huemer is mistaken about this. HI suffers from a serious defeater—the Accidentalness Objection. Adams' version of the Divine Command Theory does not suffer from that defeater. So, contrary to what Huemer says, there is a form of SR, in particular Adams' version of the Divine Command Theory, that enjoys an epistemological advantage over HI. To see this, reason as follows: According to Adams' version of the Divine Command Theory, *wrongness* and *being contrary to God's commands* are identical properties. God's commands are not causally inefficacious. For this reason it is easy to see how the Divine Command Theory can avoid the accidentalness objection: Suppose that God issues some commands. Suppose that he created human psychology so that when humans make judgments about rightness and wrongness, the pattern of those judgments matches the pattern of God's commands. God commands humans not to murder. He structures human psychology so that humans have the intuition that murder is wrong. So one form of SR, in this case Adams' Divine Command Theory, enjoys an epistemological advantage over HI. HI suffers from the accidentalness objection. This form of SR does not.

Of course, Huemer also claims that naturalist versions of SR do not enjoy an epistemological advantage over HI. He says this:

> [T]he synthetic reductionist will have to appeal to ethical intuition, just as the intuitionists do. This deprives their position of one of its central alleged advantages…. The only plausible way to maintain that we have direct awareness of moral facts would be to appeal to either 'ethical intuition' or a 'moral sense'—which is precisely the sort of 'mysterious', 'spooky' thing the reductionist wants to avoid.

Note that Richard Boyd's homeostatic consequentialism is a form of SR. And like Adams' version of the Divine Command Theory, it does not suffer from the defeater that threatens HI. Again the main point is similar to the one about Adams' theory discussed above. Goodness for Boyd is causally efficacious. So this form of SR does not suffer from the accidentalness objection.

4.7 Huemer's Argument Against SR

Huemer is not content to simply argue that SR and HI are epistemologically on a par. He also offers an argument against SR:

(1) Moral properties are radically different from the sorts of properties that proponents of SR identify with them.
(2) If two things are radically different, then they are not identical.
(3) So moral properties are not identical to the sorts of properties that proponents of SR identify with them.

The argument is valid. Premise (2) is clearly true. So the soundness of the argument hinges on whether premise (1) is true. Huemer's defense of premise (1) includes two parts. First, he claims that we intuit that moral properties are radically different from natural kind and theological properties. Second, he appeals to a principle about reference:

> [I]n my view, our having, at least roughly and for the most part, correct intuitions or beliefs about the nature of *x* is a precondition on our referring to *x*. I can talk about Plotinus without knowing much about him; about all I know is that he was a philosopher who lived a long time ago. But I could not talk about Plotinus if I had completely inaccurate beliefs (allegedly) about him; for example, if I thought 'Plotinus' referred to a cake, then I would not be talking about Plotinus (the person) at all. Similarly, if 'good' seems to us to refer to something of a fundamentally different category from natural properties, then when we say 'good' we are not talking about a natural property.

In the passage above, Huemer endorses a principle about reference. Call this principle 'Huemer's Constraint on Reference' (HCR). As I see it, the principle Huemer means to endorse is this:

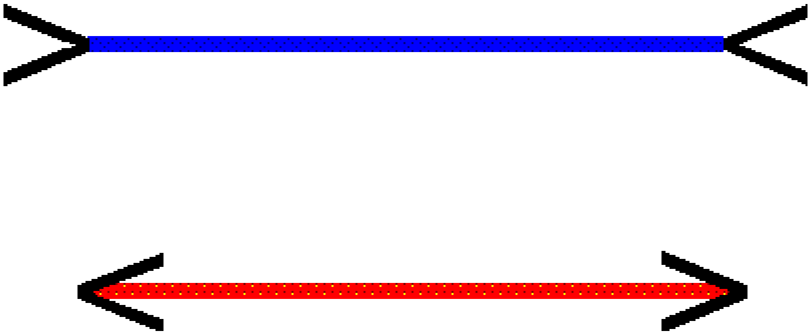
HCR: A speaker's use of 'x' refers to x only if that speaker, roughly and for the most part, has correct intuitions or beliefs about the nature of x.

Admittedly, I am doing some speculative reconstruction of Huemer's defense of (1). But I take it that his reasoning is this: We intuit that moral properties are radically distinct from natural kind and theological properties. If that intuition is false, then, roughly and for the most part, our intuitions and beliefs about the nature of moral properties are incorrect. So, given HCR, our uses of moral terms fail to refer. But our uses of moral terms succeed in referring. So our intuition that moral properties are radically distinct from natural kind and theological properties is true. So premise (1) is true.

4.8 Three Problems for Huemer's Argument Against SR

I do not find Huemer's argument on behalf of (1) to be persuasive. In what follows I undermine his argument in three ways: First, the argument relies on the claim that we intuit that moral properties are radically distinct from natural kind and theological properties. Given Huemer's account of intuitions and appearances, there are two natural readings of this claim. I argue that one of the readings might be true but it renders Huemer's argument for (1) invalid and the other reading of the relevant claim renders the argument for one valid but it is false. So whichever reading is correct, Huemer's argument is unsound. Second, I argue that an example Huemer uses to motivate (1) is not relevantly similar to the case of moral theorizing. So that example provides no support for (1). Third, Huemer's defense of (1) relies on HCR. But HCR delivers incorrect judgments about a number of examples. So HCR is false. Since Huemer's argument for (1) relies on HCR, it is unsound.

To start things off, let us go back to Huemer's theory of intuitions and appearances. In the course of stating his theory, Huemer discusses the Muller-Lyer illusion. Consider the two lines below:



Just looking at the two lines above yields the appearance that the top line is longer than the bottom line. But if one measures the two lines, they appear to be of the same length. So the two appearances conflict. In this case, the appearance yielded by way of measurement is stronger than the appearance yielded by way of just looking. So normal perceivers believe the two lines are the same length even though there is still an appearance according to which the two lines are of the same length. When more than just two appearances and beliefs are involved, things are more complicated. But, Huemer says, the general idea remains the same. We are inclined to trust the appearance that seems more strongly to be true.

The point can be made in this way: There is an *initial appearance* that the top line is longer than the bottom line. But there is also an appearance generated by measuring the two lines and an appearance generated by the intuition that measuring is a more trustworthy source of information than just looking. The result from this combination of appearances is an *overall appearance* that the two lines are of the same length. The strength of the overall appearance outweighs the strength of the initial appearance. Thus, we believe based on the overall appearance rather than the initial appearance.

In light of this, consider again one of the claims Huemer makes to motivate premise (1). He said that we intuit that moral properties are distinct from natural kind and theological properties. It seems to me that, given Huemer's account of intuitions and appearances, there are two natural ways to construe this claim:

HC1: Almost everyone has an initial intuition that moral properties are distinct from natural kind and theological properties.

HC2: Almost everyone has an overall intuition that moral properties are distinct from natural kind and theological properties.

Suppose Huemer means to be suggesting HC1. Then his argument on behalf of (1) is unsound. To see this, note that HC1 and HPC together yield no verdict whatsoever about whether our uses of moral terms successfully refers. This is because HPC puts no constraints on initial intuitions for successful reference. HPC says that a subject's use of 'x' fails to refer x if that subject's intuitions and beliefs about the nature of x are "roughly and for the most part" incorrect. But having an initial intuition that is deceiving is compatible with having intuitions and beliefs that are roughly and for the most part correct. The falsity of an initial intuition about the nature of x is compatible with the truth of an overall intuition about the nature of x. HPC only mentions overall intuitions. So if Huemer's claim is supposed to be read as HC1, then his argument is invalid.

Suppose instead that Huemer means to be suggesting HC2. Then his argument on behalf of premise (1) is unsound because HC2 is false[50]. To see this, first, grant Huemer the claim that HC1 is true. So suppose that everyone has the initial intuition that moral properties are distinct from natural kind and theological properties. While this may be true, it is false that almost everyone has the overall intuition that moral properties are distinct from natural kind and theological properties. Some people reflect on Kripke and Putnam's examples of alleged *a posteriori* necessary identities. Such reflection yields in some of these people the intuition that what is allegedly true of Kripke and Putnam's examples may be true in the cases of moral properties as well. They intuit that they have a defeater for their initial appearance that moral properties are distinct from natural kinds and theological properties. Some of these people then reflect on the fact that we have moral knowledge. When they reflect on how we could possibly have moral knowledge it seems to them that if moral properties are distinct from natural kind and theological properties, then moral knowledge is impossible. Such reflections then leave them with the intuition that moral properties are identical to natural kinds or theological properties. Combining all of these intuitions yields an overall intuition in some people that moral properties are identical to natural kind or theological properties. That overall intuition outweighs the initial intuition. So while they have the initial intuition that moral

[50] The argument of this paragraph is similar in spirit to an argument that appears in Schroeder (2009, pp. 201-2). For Huemer's response to Schroeder on this point see his (2009a, pp. 224-5). I do not think anything Huemer says there casts doubt on the way that I formulate Schroeder's point here.

properties are distinct from natural kind and theological properties, they have the overall intuition that moral properties are identical to natural kind or theological properties. What is more, it is those with this overall intuition that Huemer's argument is supposed to convince. So HC1 is true but HC2 is false. Therefore, if Huemer means to be suggesting HC2 rather than HC1, his argument for (1) is unsound since HC2 is false.

Before moving on to criticizing HCR, it will be useful to discuss how these reflections help us to see what is wrong with an example Huemer uses to illustrate how his argument is supposed to work. He says this:

> [S]uppose a philosopher proposes that the planet Neptune is Beethoven's Ninth Symphony. I think we can see that that is false, simply by virtue of our concept of Neptune and our concept of symphonies. Neptune is an entirely different kind of thing from Beethoven's Ninth Symphony. No further argument is needed. Indeed, if a person *couldn't* see that Neptune is not a symphony, we would say he either had no idea what Neptune was, or had no idea what a symphony was.

It seems to me that the discussion of HC2 just above reveals that Huemer's Neptune and the Ninth Symphony case is not relevantly similar to the case of moral properties. In the case of Neptune and the Ninth symphony, we have a strong initial intuition that Neptune is not identical to the Ninth Symphony. But, and this is the important point, we also have no appearances or beliefs of any kind whatsoever that conflict with this strong initial intuition. So what makes us so confident that Neptune is distinct from the Ninth Symphony is the presence of a strong initial intuition that the two are distinct together with the absence of any intuitions or appearances to the contrary. Things are very different in the case of moral theorizing. While many, perhaps all of us, have the initial intuition that moral properties are distinct from natural kind and theological properties, some of us have contrary intuitions and beliefs, such as the ones discussed in my argument against HC2 above. And for some of us, those contrary intuitions and beliefs are together stronger than the initial intuition. So the Neptune and Ninth Symphony case is not relevantly similar to the case at hand. It, therefore, offers no support for premise (1).

But suppose one grants Huemer the claim that HC2 is true. His motivation for premise (1) is still inadequate because HCR is false. To see this, consider the following case: Four ethicists—an ethical intuitionist, a naturalist, a theistic ethicist, and an error theorist—walk into a bar. They hope to have a quiet evening pleasantly and politely discussing the differences between them. Suddenly, fighting breaks out all around them. Things seem to be getting dangerous. At once the four ethicists look at each other and say in unison 'This is bad!' Now, no two of these ethicists have very similar intuitions or beliefs about the nature of *badness*. If one of the four has beliefs or intuitions about the nature of *badness* that are roughly and for the most part correct, then the other three have beliefs and intuitions that are not. So if HCR is true, then, the ethicists do not refer to the same thing when they use 'bad'. But they do refer to the same thing when they use 'bad'. So HCR is false.

Here is another case—this time inspired by Huemer's 'Plotinus' example: Two historians walk into a bar. Historian$_1$ believes that Plotinus is a person. Historian$_2$ believes that Plotinus is a cake. They are familiar with the same body of literature and the same pro-person and pro-cake arguments. They spend the evening defending their respective positions. Historian$_1$'s intuitions and beliefs about the nature of Plotinus are roughly and for the most part correct. Historian$_2$'s intuitions and beliefs are not. After they each have had too many beers, things get heated. Historian$_1$ looks at his interlocutor, pounds on the table, and yells "Plotinus was a person!" Historian$_2$ stares back angrily and yells "Plotinus was a cake!" If HCR is true, then Historian$_1$ and

Historian$_2$'s uses of 'Plotinus' do not refer to the same thing. But their uses of 'Plotinus' do refer to the same thing. So it seems to me that Huemer's 'Plotinus' case is underdescribed. And certain ways of filling out his case show that HCR is false[51].

Here is a third case: Thales walks into a bar and starts drinking. To his left are four ethicists. To his right are two historians. In the corner is an upset looking robot. After one too many beers Thales becomes overconfident, stands on a bar stool, and starts preaching to the other patrons about his deep conviction that water is the arche of all things. He passionately explains that earth, air, and fire, for example, are all fundamentally water. This disturbance angers many of the other patrons. Fighting ensues. The historians join in on the fighting. The ethicists shout something in unison and then run for the door. The robot just stays put looking downcast. Thales' beliefs about the nature of earth, air, and fire are, roughly and for the most part, incorrect. So if HCR is true, then Thales' uses of 'earth', 'air', and 'fire' fail to refer. But Thales uses of these terms do not fail to refer. So HCR is false.

These three cases reveal another problem with Huemer's argument. Huemer defends premise (1) by appealing to HCR. But HCR is too rigid. It excludes numerous cases of successful reference. So HCR is false. Huemer's defense of (1) is therefore unsuccessful. He needs to find some other motivation for (1).

[51] There is an actual case like this. British philologist John M. Allegro's book *The Sacred Mushroom and the Cross* argues that the historical Jesus was not a person but instead a hallucinogenic mushroom!

CHAPTER 5

INTUITION, EVIDENCE, AND A PRIORITY

In a series of interesting, important, and widely cited papers, George Bealer has defended a theory of a priori knowledge. In this paper I identify two problems for his theory. One problem is that it is impossible for anyone to satisfy the conditions Bealer's theory places on a priori knowledge. The other problem is that, although Bealer's theory is supposed to explain why intuition is a basic source of evidence, it instead implies that intuition is not a basic source of evidence. The first problem can be resisted, to a certain extent, by introducing some revisions. The second problem is much harder to resist.

5.1 Why Trust Intuition?

It is standard practice in philosophy to construct theories on the basis of intuitions. But why should philosophers trust their intuitions? George Bealer's ((1987), (1996a), (1996b), (1998a), (1998b), (1998c), (1999), and (2000)) answer is that intuition is a basic source of evidence. The reason intuition is a basic source of evidence, he says, is that it enjoys a strong modal tie to the truth. Bealer has offered an intricate theory that is supposed to explain why this tie between intuition and truth obtains[52].

5.2 Bealer's Theory

Intuition: Bealer's theory includes an account of intuition. Intuition is a particular sort of seeming. A seeming can be perceptual, introspective, or intellectual. An intellectual seeming is an intuition. A perceptual or introspective seeming is not an intuition. Intuitions are distinct from beliefs. One can believe that a proposition is true without intuiting that it is true. One can intuit that a proposition is true without believing

[52] Other recent criticisms of Bealer appear in Beebe (manuscript), Devitt (2011), Jenkins (2008), Sarch (2010), and Schechter (2010). A number of older critical discussions of Bealer's theory, together with his replies, appear in the same volumes as Bealer (1996a), (1998a) and (1999).

that it is true. Intuitions are fallible. One can intuit that a proposition is true even if that proposition is false. Some intuitions are a priori. To intuit a proposition a priori is to intuit that it is necessarily true. Bealer's project concerns strictly a priori intuitions. Examples of intuitions, in Bealer's sense, are the intellectual seeming that '2 + 2 = 4' is necessarily true and the intellectual seeming that 'torturing babies just for fun is wrong' is necessarily true.

Cognitive Conditions: Bealer (1999, p. 39-40) identifies three sorts of cognitive condition—intelligence, attentiveness, and memory. A subject is in good cognitive conditions if that subject is to some significant degree intelligent, attentive, and possessed of good memory.

Conceptual Repertories: Bealer discusses conceptual repertories. Suppose that a subject introduces the term 'chromic' and that 'chromic' expresses *the property of being chromic*. Suppose also that she applies 'chromic' to phenomenal color (red, blue, purple, etc.) and she refrains from applying 'chromic' to phenomenal black or phenomenal white. Suppose further that she has never experienced phenomenal gray, she cannot yet conceive of phenomenal gray, that the question "Is phenomenal gray chromic?" has never occurred to her, and that she has no intuitions about whether phenomenal gray is chromic. Now suppose that the subject experiences phenomenal gray for the first time. As a result, she thinks of new questions relevant to the extension of 'chromic' and she has new intuitions about which colors are chromic and which are not. Suppose, for example, that she has the intuition that phenomenal gray is not chromic. It is in the context of this example that Bealer (1999, p.40) offers a concrete example of a conceptual repertory:

> [Before the subject experiences phenomenal grey], the decisive cases involve items… which lie beyond her experience and conceptual repertory. [S]he cannot even entertain the relevant test questions, let alone have truth-tracking intuitions regarding them…. There is no requirement

> that… a person must already have experiential and/or conceptual resources sufficient for deciding the possible extensions of the concept.

Consider the subject prior to her experience of phenomenal gray. She has a conceptual repertory that included questions such as “Is phenomenal black chromic?” and “Is phenomenal purple chromic?” and so on. Furthermore, her conceptual repertory included various intuitions about the property of being chromic such as the intuition that phenomenal black is not chromic and the intuition that phenomenal purple is chromic and so on. Next consider the subject after her experience of phenomenal gray. Her conceptual repertory expanded. Now it includes additional questions such as “Is phenomenal gray chromic?” and additional intuitions such as the intuition that phenomenal gray is not chromic. The subject’s experience of phenomenal grey caused her conceptual repertory to expand. This suggests the following account of conceptual repertoires: A subject’s conceptual repertory with respect to a property, P, is all the questions the subject can entertain and all the intuitions the subject has about what things instantiate P and what things fail to instantiate P.

5.2.1 Intuition as Basic Evidence

Bealer distinguishes between basic and non-basic evidence. Intuition and phenomenal experience are basic sources of evidence. Testimony is a non-basic source of evidence. Bealer argues that in order for a candidate source of evidence to count as basic, it must have a Strong Modal Tie to the Truth.

5.2.2 Strong Modal Tie to the Truth

Different passages suggest different accounts of the sort of tie to the truth that Bealer is interested in. Consider this typical passage (1999, p. 35-36):

> We are left with modal reliabilism, according to which something counts as a basic source of evidence iff there is an appropriate kind of strong modal tie between its deliverances and the

> truth…. This suggests an analysis along the following lines: a candidate source is basic iff for cognitive conditions of some suitably high quality, necessarily, if someone in those cognitive conditions were to process theoretically the deliverances of the candidate source, the resulting theory would provide a correct assessment as to the truth or falsity of most of those deliverances.

Passages such as this suggest that a candidate source of evidence has a strong modal tie to the truth if and only if, necessarily, if a subject that is in good cognitive conditions forms beliefs on the basis of a systemization of the deliverances of that source of evidence, then most of those beliefs are true.

Other passages, however, suggest that Bealer (1996, p. 140, 139) means something stronger. Consider, for example, these passages:

> [T]his procedure is an idealization…. These efforts are typically collective, and the results of past efforts—including those of past generations—are used liberally. The fact that speech and writings are used does not disqualify these collective efforts as *a priori*, at least not according to the central use of '*a priori*' I am employing. Experience and/or observation can be used to raise—and also to resolve—doubts about the quality of the communication conditions (speaker and author sincerity, reliability of the medium of transmission, accuracy of interpretation, etc.). But these empirical resources play no role in the procedure of *a priori* justification itself.
>
> It is of course another matter whether it is nomologically possible for human beings to be in sufficiently good cognitive conditions to achieve the kind of autonomy and authority asserted as a mere possibility in these two theses. Whether this is nomologically possible is a question on which I take no stand here. My personal belief, however, is that collectively, over historical time, undertaking philosophy as a civilization-wide project, we can do so closely enough to obtain authoritative answers to a substantial number of central philosophical questions.

This suggests a more holistic or collective account of the sort of tie to the truth in question: Suppose there is a large group of people. Suppose that most of these people are in good cognitive conditions. Suppose, in addition, that these people repeatedly consult a candidate source of evidence and repeatedly systematize its deliverances. They do not do this as individuals. They get together and talk about the relevant deliverances and about how to systematize them. They do not do this a few times. They do it many times. I will call any group like this a "Talkative Group". If a Talkative Group uses a source of evidence, E, in the way just described, then I will say that "the Talkative Group systematizes E while in good cognitive conditions." When talking about the beliefs of members of a Talkative Group that are formed as a result of a process like this, I will say that "the Talkative Group's beliefs are formed on the basis of E." A candidate source of evidence, E, therefore, enjoys a Strong Modal Tie to the Truth if and only if, E satisfies the following condition:

(SMTT): Necessarily, any Talkative Group that systematizes E while in good cognitive conditions is such that most of the beliefs of most of the members of that Talkative Group that are formed on the basis of E are true.

5.2.3 Determinate Understanding

TPIs: Bealer discusses test property identities (TPIs). A TPI is a sentence of the form 'the property of being F = the property of being A' where A is some formula. Examples of TPIs include 'The property of being wrong = the property of failing to maximize utility' and 'The property of being a triangle = the property of being a three sided polygon'.

Ways of Understanding: Bealer discusses ways of understanding. Concrete examples and precise conditions identifying the ways of understanding Bealer has in mind are hard to find. But there are a few conditions constraining what ways of understanding can be that Bealer explicitly endorses and a few others that are implied by his theory.

First, Bealer (1999, p. 42) requires that ways of understanding are natural, non-ad-hoc, and non-Cambridge-like. For example, he says that the "intention here is that [the term 'way of understanding'] ranges over *natural* modes of understanding (i.e., non-*ad-hoc* modes of understanding)." Thus we have our first constraint on ways of understanding:

Naturalness (NAT): Ways of understanding are natural.

Second, further constraints appear in Bealer's discussion of the chromic example. Bealer considers two variants of the example. In the first variant, the woman has until a certain time never experienced phenomenal gray. Then she experiences it. Bealer insists that after experiencing phenomenal gray the woman continues to understand chromic in the same way that she did prior to experiencing phenomenal gray. In the second variant, it is nomologically impossible for the woman to experience phenomenal gray. But she has a counterpart who is capable of experiencing phenomenal gray. Again, Bealer insists says that the woman understands chromic in the same way as her counterpart and in the same way as the woman in the first example. In each case, the woman or her counterpart receives an expanded conceptual repertory. But she continues to understand chromic in the same way. Similar points apply to cases in which the woman's cognitive conditions improve. Bealer (1999, p. 40) says it this way:

> There is no requirement that, in order to possess a concept determinately, a person must already have experiential and/or conceptual resources sufficient for deciding the possible extensions of the concept….This could be so without there being any (immediate) shift in the way the woman (or her counterpart) understands any of her concepts or the propositions involving them…. Of course, the same sort of thing could happen in connection with nomologically necessary limitations on aspects of the woman's cognitive conditions (intelligence, attentiveness, memory, constancy, etc.)

So we have two more constraints on ways of understanding:

Survival of Improved Cognitive Conditions (SIC): Improving a subject's cognitive conditions does not, by itself, change the way the subject understands a TPI.

Survival of Expanded Conceptual Repertory (SEC): Expanding a subject's conceptual repertory does not, by itself, change the way the subject understands a TPI.

As far as I can tell, these are the only constraints on ways of understanding that Bealer discusses. However, I am going to attribute to Bealer two further constraints. Although he does not discuss them, they are implied by his theory. One of the constraints is needed to keep a central condition included in one of his definitions from being trivial. Both constraints are needed to avoid a counterexample I discuss below. I will offer explicit motivations for these attributions after I have discussed Bealer's theory in greater detail. Where the relevant details are discussed, motivation for these attributions will appear in footnotes[53]. For now I will just state the conditions in question:

Survival of Change in Truth-Value (SCT): Changing a subject's belief about the truth-value of a TPI does not, by itself, change the way in which the subject understands that TPI.

Survival of Empirical Revelation (SER): Revealing to a subject empirical information about how that subject arrived at his or her belief about the truth-value of a TPI does not, by itself, change the way in which that subject understands that TPI.

A Priori Stability: Bealer defines a priori stability in a very technical and formalistic way. I am going to try to translate his definition into much simpler and plainer language. Here is the way Bealer (1999, p. 42)[54] states the definition:

> Consider an arbitrary property-identity p which someone x understands m-ly. Then, x settles with a priori stability that p is true iff, for cognitive conditions of some level *l* and for some conceptual repertory *c*, (1) x has cognitive conditions of level *l* and conceptual repertory *c* and x attempts to elicit intuitions bearing on p and x seeks a theoretical systemization based on those intuitions and that systematization affirms that p is true and all the while x understands p m-ly, and (2) necessarily, for cognitive conditions of any level *l'* at least as great as *l* and for any conceptual repertory *c'* which includes *c*, if x has cognitive conditions of level *l'* and conceptual repertory *c'* and x attempts to elicit intuitions bearing on p and seeks a theoretical systematization based on those intuitions and all the while x understands p m-ly, then that systematization also affirms that p is true.
>
> The idea is that, after x achieves <*c*, *l*>, theoretical systematizations of x's intuitions always yield the same verdict on p as long as p continues to be understood m-ly throughout. That is, as long as p is understood m-ly, p always gets settled the same way throughout the region to the "northeast" of <*c*, *l*>.

With this passage in mind, I will attribute to Bealer the following account of a priori stability: Suppose that a subject is in good cognitive conditions and has a reasonably large conceptual repertory. Suppose, further, that the subject judges some TPI, P, to be true. Next, suppose the subject arrived at this judgment by, first, eliciting intuitions about P, then systematizing those intuitions, and then judging, on the basis of that systemization, that P is true. Suppose, additionally, that the subject understands P in the

[53] See footnotes 6 and 7.

[54] Essentially identical definitions appear in Bealer (1998a, pp. 286-7) and (2000, p. 16).

same way throughout this process. Finally, suppose that as long as the subject continues to understand P in this way, she satisfies the following condition:

> (APS): Necessarily, for any improvement in the subject's cognitive conditions, for any expansion of the subject's conceptual repertory, and for any attempt by the subject to again elicit intuitions about P and to again systematize those intuitions, she will, on the basis of the new systemization, again judge that P is true.

The relevant subject has settled on P with a priori stability if and only if each of these suppositions obtain.

In general, for any subject, S, and any TPI, P, S settles on P with a priori stability if and only if S is in good cognitive conditions, S has a reasonably large conceptual repertory, S elicits and systematizes intuitions about P, S judges that P is true on the basis of that systemization, S understands P in the same way throughout this process, and S satisfies (APS) as long as S continues to understand P in that way[55].

Determinate Understanding: It is now possible to state the final component of Bealer's theory. According to Bealer, there is a special way of understanding TPIs that allows for a priori knowledge. This way of understanding is determinate understanding. A subject, S, that determinately understands a TPI, P, is a subject who understands P in a special way. Any other subject who understands any other TPI in the way that S understands P is such that that other TPI is true if and only if it is possible for that other subject to settle on that other TPI with a priori stability. It can be said this way:

[55] Now we are in a position to motivate my attribution of (SCT) to Bealer. If Bealer denies (SCT), then all possible subjects trivially satisfy (APS). Any subject that changes her mind about a TPI after becoming smarter will no longer understand that TPI in the same way. So it is trivially true that as long as the subject understands the relevant TPI in the same way, then any improvements in her cognitive conditions or expansions of her conceptual repertory will yield a systemization with the same verdict about that TPI. So every subject trivially satisfies (APS). And there is therefore no point in adding (APS) to the definition of a priori stability. But it is implausible to suggest that Bealer intended to add a trivial, pointless condition to his definition. So Bealer must accept (SCT) even though he never explicitly discusses it.

(DU): A subject, S, determinately understands a TPI, P, if and only if S understands P in the following way: for any subject, S*, and any TPI, P*, such that S* understands P* in the way that S understands P, P* is true if and only if it is possible for S* to settle on P* with a priori stability.

Intuition as a Basic Source of Evidence

Bealer claims in several places that his theory explains why intuition is a basic source of evidence. Just after he finishes his discussion of determinate understanding, for example, he (1999, p. 47) says:

> In the course of our discussion of the evidential force of intuitions, we noted a shortcoming in traditional empiricism and traditional rationalism, namely, that neither successfully explains why intuition and phenomenal experience should be basic sources of evidence. Modal reliabilism filled this explanatory gap: the explanation is that these two sources of evidence have the right sort of modal tie to the truth. We saw, moreover, that neither traditional empiricism nor traditional rationalism successfully explains why there should be such a tie between these basic sources and the truth. The analysis of determinate concept possession fills this gap: In the case of intuition, determinate possession of our concepts entails that there must be such a tie.

While it is clear that Bealer thinks his theory implies that intuition satisfies (SMTT), he says very little about how to establish this implication. For this reason, I am going to have to offer some speculative reconstructions of his argument. Here are two plausible candidate arguments:

Argument B1

(1) If a Talkative Group, G, forms beliefs about some area of inquiry, A, on the basis of intuition, G's members are in good cognitive conditions, and G's members determinately understand the TPIs within A, then most beliefs about A held by most of G's members are true.
(2) If (1), then intuition satisfies (SMTT).
(3) If intuition satisfies (SMTT), then intuition is a basic source of evidence.
(4) Therefore, intuition is a basic source of evidence.

Argument B2

(1) If a Talkative Group, G, forms beliefs about some area of inquiry, A, on the basis of intuition and G's members are in good cognitive conditions, then G's members determinately understand the TPIs within A.
(2) If (1), then it is necessary that most beliefs about A held by most of G's members are true.

(3) If it is necessary that most beliefs about A held by most of G's members are true, then intuition satisfies (SMTT).
(4) If intuition satisfies (SMTT), then intuition is a basic source of evidence.
(5) Therefore, intuition is a basic source of evidence.

5.3 The Problem of Absent Determinate Understanding

Bealer claims that it is possible for some subjects to determinately understand some TPIs. He (1999, p. 38) says, for example, this:

> Now, intuitively, it is at least possible for most of the central concepts of the a priori disciplines to be possessed determinately by some cognitive agent or other (e.g., such concepts as conjunction, negation, identity, necessity, truth, addition, multiplication, set membership, quality, quantity, relation, proposition, consciousness, sensation, evidence, justification, knowledge, explanation, causation, goodness, etc.). It would be quite ad hoc to deny this.

In other places, Bealer (1999, p. 48, 36) makes the even stronger claim that not only is it possible for some subject to determinately understand some TPIs, but that actual humans in some cases "approximate" such understanding.

> [I posit] only the metaphysical possibility of autonomous a priori knowledge, perhaps on the part of creatures in cognitive conditions superior to ours. But, if true, the thesis would nevertheless help to illuminate our own situation. For to the extent that we *approximate* the indicated cognitive conditions, we are able to *approximate* the sort of autonomous a priori knowledge contemplated in the thesis.
>
> [W]hen we limit ourselves to suitably elementary propositions, then relative to them we *approximate* such cognitive conditions. For suitably elementary propositions, therefore, deliverances of our basic sources would provide in an approximate way the kind of pathway to the truth they would have generally in the envisaged high-level conditions.

Of course, in the second of these passages, Bealer doesn't explicitly mention determinate understanding. But he does mention a priori knowledge. And, as we have seen, to possess a priori knowledge of a TPI, an individual, or a community's members, must determinately understand that TPI. So the second passage entails that humans in some sense "approximate" determinate understanding of some TPIs.

This claim is needed to motivate Argument B1 and Argument B2. First, consider Argument B1. If it is impossible for any subject to determinately understand any TPI, then premise (1) will be a counterpossible. If counterpossibles are trivially true, then (1) will be true. If counterpossibles have non-trivial truth values, then perhaps (1) will still be true. But it is difficult to see what the motivation for premise (2) of Argument B1 could be. Why think that intuition enjoys a strong modal tie to the truth just because some subjects that understand something in a metaphysically impossible way would be guaranteed to have true beliefs if they relied on intuition? Without the claim that some subjects determinately understand some TPIs, premise (2) is without motivation.

Next consider Argument B2. If it is impossible for any subject to determinately understand any TPI, then premise (1) is false. Premise (1) says that if a Talkative Group forms beliefs about some TPIs on the basis of intuition, then the members of that Talkative Group determinately understand those TPIs. But if no one can determinately understand any TPI, then forming beliefs about some TPIs on the basis of intuition does not imply determinate understanding. So if it is impossible for any subject to determinately understand any TPI, then premise (1) of Argument B2 is false.

So in order for Bealer's theory to have the implication he advertises, it needs to be coupled with the substantive claim that some subjects, and perhaps some actual humans, determinately understand some TPIs. The problem is, this substantive claim is false. It is impossible for any subject, other than perhaps God, to determinately understand any TPI. To see why, it will be useful to start with the case just below.

5.3.1 ROBOTRON A and ROBOTRON B

Suppose that Black knows which moral TPIs are true and which are false. Suppose, further, that he builds two robots—ROBOTRON A and ROBOTRON B. Black programs ROBOTRON A and ROBOTRON B so that all true moral TPIs seem true to them and all false moral TPIs seem false. He builds the two robots so that, to the same degree, they are in good cognitive conditions. They are both very intelligent. Neither is smarter than the other. They both are as attentive as a robot can possibly be. They both have flawless memories. In addition, the two robots have identical, extremely large, conceptual repertories. So there is no question about any of these TPIs that one of the robots can entertain and that the other robot cannot. And there is no intuition about the answer to one of those questions that one of the robots has that the other robot lacks.

The only relevant differences between ROBOTRON A and ROBOTRON B are these: Black loves ROBOTRON A and he hates ROBOTRON B. He expresses his love and his hate by placing certain limits on ROBOTRON B’s mental abilities. ROBOTRON B was built with an intelligence monitor. If his intelligence ever increases by 10^1000 degrees, the monitor will trigger a digital recording of Black explaining that ROBOTRON B’s moral intuitions were determined by the output of a Random Truth-Value Assigner (RTA). ROBOTRON B loves Black and trusts him. He has every reason to think that Black is a trustworthy source of information. So if someone upgrades ROBOTRON B’s intelligence by 10^1000 degrees and that digital recording plays, ROBOTRON B will change his mind about the truth-values of the relevant moral TPIs. He will become agnostic about those truth-values or perhaps even believe that they are false. So if he becomes intelligent enough, he will lose a lot of true beliefs about the relevant TPIs. And if he keeps those true beliefs, ROBOTRON B will remain at a limited level of intelligence. ROBOTRON A, however, was not built with such a monitor. And Black explained to him the difference between the two robots and the motivation for building them differently and he directed ROBOTRON A to never tell ROBOTRON B about any of this. So no matter how much smarter ROBOTRON A gets, he will keep his true beliefs about the TPIs in question. He is not limited in the way that ROBOTRON B is limited.

Now, a little reflection on this story reveals that ROBOTRON B does not determinately understand any of the relevant TPIs. A little further reflection reveals that ROBOTRON A, for the same reasons, does not determinately understand any of the TPIs in question either. A little additional reflection reveals that, for the exact same reasons, no actual human nor any possible subject (other than God perhaps) determinately understands any TPI.

Start with ROBOTRON B. He does not determinately understand any of the relevant moral TPIs. To see this, pick any of the relevant TPIs that you want, say P. Because of the digital recording, ROBOTRON B would change his mind about P if he were made more intelligent. So it is possible for ROBOTRON B to change his mind by improving his cognitive conditions. And, since it is possible that such improvements will change his mind, it is necessarily possible. So, necessarily, it is not necessary that such improvements will not change ROBOTRON B's mind. So it is not possible that, necessarily, such improvements will not change his mind about P. So it is not possible for ROBOTRON B to satisfy (APS). Now, remember that P is true. So P is true even though it is not possible for ROBOTRON B to settle on P with a priori stability. So it is not the case that P is true only if it is possible for ROBOTRON B to settle on P with a priori stability. So (DU) is unsatisfied. ROBOTRON B does not determinately understand P or any of the other moral TPIs.

ROBOTRON A does not determinately understand any of the relevant TPI's either. Like ROBOTRON B, it is impossible for ROBOTRON A to satisfy (APS).

Perhaps, as things actually are, it is very unlikely that ROBOTRON A would change his mind about the relevant TPIs after receiving improved cognitive conditions. But it is possible that he would change his mind. Black might have built him in the same way that he built ROBOTRON B. Black might implant such a device tomorrow. So it is possible that an increase in the degree to which ROBOTRON A is in good cognitive conditions would get him to change his mind about the relevant TPIs. And, just as in the case of ROBOTRON B, these possibilities, however remote in ROBOTRON A's case, are all that is needed to establish that it is impossible for ROBOTRON A to satisfy (APS) and, therefore, impossible for him to determinately understand any of the relevant moral TPIs.

No actual human determinately understands any TPIs either. Pick any actual human you want and any TPI that that human believes. That human does not satisfy (APS) with respect to any of those TPIs. It might have turned out that, as a quirk of evolution, once that human reaches a certain level of intelligence or remembers a certain thing or pays enough attention, she will change her mind. It might have turned out that aliens implanted a Blackesque device in her brain or stand ready to implant such a device tomorrow. It might turn out that a quantum fluctuation will materialize such a device in that human's brain next week. So it is possible that the human would change her mind after receiving improved cognitive conditions. So it is impossible for any human to satisfy (APS) with respect to any TPI and therefore impossible for any human to determinately understand any TPI.

Similar considerations show that no possible subject, with perhaps the exception of God, can determinately understand any TPI. God might get off the hook because, being omniscient, he can't help but have true beliefs about each of those TPIs and, necessarily, he won't change his mind about any of them.

5.3.2 No New Empirical Information

Someone might think there is a simple way to fix Bealer's theory. All of the possibilities I have highlighted are cases in which an increase in the subject's intelligence will trigger, as a result of tampering, the revelation of new empirical information relevant to the truth-value of the TPIs in question. In the case of ROBOTRON B, for example, Black arranged things so that certain improvements in ROBOTRON B's cognitive conditions will trigger the revelation of new empirical information about the causal origins of his beliefs relevant to the truth-value of P. A simple modification to (APS) will fix this problem. Just exclude from consideration improvements in a subject's cognitive conditions that bring about the revelation of new empirical information relevant to the truth-value of P:

(APS*): Necessarily, for any improvement in S's cognitive conditions, for any expansion of S's conceptual repertory, and for any attempt by S to again elicit intuitions and to again systematize those intuitions *that is not accompanied by new empirical information relevant to the truth-value of P*, S will again judge on the basis of that systemization that P is true.

This modification handles the case of ROBOTRON B. If ROBOTRON B's intelligence were improved by 10^1000 degrees, new empirical information about the origins of his beliefs would be revealed to him. While these considerations preclude ROBOTRON B from satisfying (APS), they are irrelevant to the satisfaction of (APS*). So ROBOTRON B, it would seem, does satisfy (APS*) and therefore the revised definition of a priori stability. Similar remarks apply to ROBOTRON A, to actual humans, and to all other possible subjects.

There is some textual evidence that Bealer would not like the fix I have proposed. In his earliest work on this topic he suggests that the procedure he identifies does not rely substantively on empirical investigation. But he then qualifies (1996, p. 140) this by saying that there are some non-substantive senses in which empirical investigation is relevant:

> Empirical beliefs—and the experiences and observations upon which they are based—are sometimes used to raise and to resolve doubts about the quality of the background cognitive conditions (intelligence, etc.)…. When I speak of not needing to rely *substantively* on empirical science, this is one of the points I have in mind.

So perhaps Bealer would not like (APS*). Regardless, adding it enables Bealer's theory to avoid the problem I have identified. So it is worth considering whether or not it can be made to work.

5.3.3 ROBOTRON C

Unfortunately, this modification is not quite right. It makes it too easy to satisfy the account of a priori stability. Suppose that Smith builds a robot—ROBOTRON C. Suppose that Smith programs all moral TPIs into a Randomized Truth Value Assigner (RTA). Now, as a matter of sheer luck, the RTA assigns 'true' to each true moral TPI and 'false' to each false moral TPI. Smith then programs ROBTRON C in such a way that his moral intuitions match the deliverances of the RTA. Each true moral TPI seems true to ROBOTRON C. Each false moral TPI seems false to him.

Smith builds ROBOTRON C so that he is in good cognitive conditions. His memory is almost perfect. He is missing just one memory. Once Smith told ROBOTRON C that his moral intuitions were formed on the basis of the RTA's deliverances. But ROBOTRON C forgot. He is very attentive. But if he had paid just a little more attention, ROBOTRON C would have picked up on subtle cues Smith had given him indicating that his moral convictions were formed on the basis of the RTA's deliverances. ROBOTRON C is very intelligent. But he is not quite smart enough to solve a series of riddles Smith had posed to him. The solution to each of those riddles is that his moral convictions were formed on the basis of the RTA's deliverances.

Smith built ROBOTRON C so that he has a huge conceptual repertory. He can think of as many questions relevant to those TPIs as any being could. If there are a finite number of relevant questions, then he has thought of each one. Furthermore, he has strong intuitions about the answers to each of those questions.

Now consider any of the moral TPIs ROBOTRON C believes, say P. If his memory were improved just a little, and he remembered what Smith said, ROBOTRON C would resystematize his intuitions and no longer judge that P is true. If he were even slightly more attentive, he would have picked up on Smith's subtle cues, resystematized his intuitions, and changed his mind about P. If he were just a little smarter, he would solve Smith's riddles, resystematize in light of them, and go agnostic about P[56].

Notice that each of these improvements in ROBOTRON C's cognitive conditions would unearth new empirical information relevant to the truth-value of P. The information advises ROBOTRON C to abandon his correct judgment that P is true. Since

[56] This case illustrates the need to attribute (SCT) and (SER) to Bealer. If Bealer does not accept these constraints on ways of understanding, then ROBOTRON C satisfies (APS) as well as (APS*) and this case is therefore a counterexample to the original, and not just the revised, definition of a priori stability.

To see this, reason as follows: Suppose that Bealer rejects (SCT). So changing a subject's belief about the truth-value of a TPI changes the way in which that subject understands that TPI. Then revealing to ROBOTRON C the origins of his intuitions will change his belief about the truth-value of P. So that will change the way in which he understands P. So it will still be true that as long as he understands P in the same way, he won't change his mind about P. So he will still have settled on P with a priori stability.

Now consider (SER). Suppose it is false. So that revealing to a subject empirical information about the origins of her beliefs about the truth-value of a TPI changes the way in which that subject understands that TPI. Then revealing to ROBOTRON C the origins of his intuitions will change the way in which he understands P. So, again, he will have settled on P with a priori stability since it is still true that as long as you don't change the way in which he understands P, he won't change his mind about P. So Bealer must accept (SER).

such information is new and empirical, it is irrelevant to whether ROBOTRON C satisfies (APS*). So ROBOTRON C has settled with a priori stability that P is true. It is therefore possible that ROBOTRON C settles with a priori stability that P is true.

Now we can make a case for the claim that ROBOTRON C determinately understands P. After all, P is necessarily true. So there are no metaphysically possible cases in which it is possible for ROBOTRON C to settle with a priori stability that P is true but P is false. And since it is possible for ROBOTRON C to settle on P with a priori stability, it is necessarily possible. So there are no metaphysically possible cases in which P is true but it isn't possible for ROBOTRON C to settle on P with a priori stability. Thus, P is true if and only if it is possible for ROBOTRON C to settle with a priori stability that P is true. ROBOTRON C, it would seem, determinately understands P. But he does not know that P is true. So something is wrong with the account of determinate understanding.

Now imagine an entire population of robots like ROBOTRON C that determinately understand numerous moral TPIs in the same way that ROBOTRON C determinately understands P. These robots form a Talkative Group. Bealer's theory implies that these robots possess a priori knowledge of morality. But they do not.

5.3.4 No New Empirical Information

One might try another fix. In addition to (APS*) and the other conditions mentioned in the present formulation of a priori stability, one might add the following condition:

(EMP): There is no empirical information relevant to the truth-value of P of which S is unaware and of which S would become aware if S's cognitive conditions were improved or S's conceptual repertory were expanded.

This way of formulating the definition of a priori stability seems to handle my counterexamples. In the case of ROBOTRON A and each actual human, improving that robot or that human's cognitive conditions *might* reveal to him or her new, empirical information relevant to the truth-value of P. There is a remote possibility that such improvements will reveal to them such information. So they fail to satisfy (APS) but succeed in satisfying (APS*). And yet, because the relevant possibilities are so remote, it is not the case that such improvements *would* reveal to them new empirical information relevant to the truth-value of P. So they succeed in satisfying (EMP). Thus, ROBOTRON A, and perhaps many actual humans, succeed in settling on some TPIs with a priori stability. The revised definition also handles ROBOTRON C. He succeeds in satisfying (APS*) for the reasons discussed above. But he fails to satisfy (EMP). In his case, the relevant possibilities are not remote at all. The slightest improvement in his cognitive conditions would definitely reveal to him new empirical information relevant to the truth-value of P. So ROBOTRON C does not satisfy (EMP). So we have a way of formulating Bealer's theory that accommodates the claim that ROBOTRON A, and perhaps some actual humans, determinately understand some TPIs without committing Bealer to the claim that ROBOTRON C determinately understands P.

As is the case with the previous revision, it is doubtful that Bealer would be happy with (EMP). The problem is that Bealer introduces the account of a priori stability in order to avoid formulating his theory in terms of 'would'[57]. Rejecting an earlier formulation of determinate understanding, he says this:

> A problem with this analysis is that it relies on the subjunctive 'would', but there are well-known general objections to relying on subjunctives in settings such as this. The solution is to replace the subjunctives with a certain ordinary modal notion. I will call this modal notion *a priori stability*.

[57] See Bealer (1998a, p. 283) and (1999, p. 41-2).

However, adding (EMP) to the definition of a priori stability fixes the problem that concerns us here. And the problem I discus below targets the revision and the unrevised theory alike.

5.4 The Problem of Lack of Basicness

This process of refining leads us to what I think is the deepest problem for Bealer's theory. In what follows, I will argue that premise (2) of Argument B1 and premise (1) of Argument B2 are false. Intuition does not satisfy (SMTT). So Bealer's theory implies that intuition is not a basic source of evidence. To see why, it will be useful to start with the case below.

5.4.1 ROBOTRON Zero

Suppose that Jones builds a robot—ROBOTRON Zero. Suppose further that Jones programs all moral TPIs into an RTA. The RTA then goes to work and randomly assigns truth-values to each of those moral TPIs. As it so happens, the result is a consistent set of moral propositions. There are no contradictions. But in many cases the truth-values assigned by the RTA do not match the actual truth-values of the relevant moral TPIs. Many true TPIs are assigned 'false' and many false TPIs are assigned 'true'. Now, suppose that Jones programs ROBOTRON Zero so that which moral TPIs seem true to him and which moral TPIs seem false to him match exactly the deliverances of the RTA.

Suppose also that Jones builds ROBOTRON Zero so that he is in good cognitive conditions. He is very intelligent. There are a few puzzles he isn't quite smart enough to solve. But none of those puzzles has as its solution any information relevant to the causal origins of his beliefs. There are not any inconsistencies in his beliefs about the relevant

TPIs. He didn't commit any fallacies. So becoming smarter will not unearth any inconsistencies or fallacies of which he is now unaware. He is very attentive. There are a few subtle cues Jones has given him that he hasn't picked up on. But the correct reading of those cues contains no new information relevant to the causal and historical origins of his beliefs. At a few formal parties, Jones tried to subtly suggest to ROBOTRON Zero that he was drinking too much. But ROBOTRON Zero just didn't pick up on the hints. ROBOTRON Zero has an almost flawless memory. The only thing he has ever forgotten is a fishing trip he took with Jones last week. However, no information about the causal and historical origins of his beliefs occurred on that trip. It was nothing more than some pleasant fishing and some pleasant small talk about Britney Spears' upcoming album.

Suppose, finally, that Jones built ROBOTRON Zero so that he has a very large conceptual repertory. He can ask as many relevant questions about those property identities as any robot possibly could. And he has strong intuitions about the answers to each of those questions.

Now, pick any of the false moral TPIs that ROBOTRON Zero believes, say P. ROBOTRON Zero has settled on P with a priori stability. He is in good cognitive conditions. He has elicited intuitions about P, systematized them, and judged on the basis of that systemization that P is true. There is no empirical information relevant to the truth-value of P that would be revealed to him if his cognitive conditions were improved. He would only remember a conversation about Britney Spears' upcoming album from a recent fishing trip, solve puzzles irrelevant to P, or feel embarrassed because Jones thought he was drinking too much at a party. None of that is relevant to P. So he satisfies (EMP). Moreover, he satisfies (APS*). Any improvement in ROBOTRON Zero's cognitive conditions that is not accompanied by new empirical information relevant to P will not yield a change in his judgment about P. How could it? Consider an improvement in his memory. The only thing that can be improved is that he would remember the fishing trip with Smith. But that fishing trip contains no information relevant to the origins of his beliefs or even any information about P at all. Why would adding a memory of the fishing trip, that has nothing to do with P, produce new intuitions in ROBOTRON Zero or cause him to systemize his intuitions differently? How could just remembering a fishing trip and a conversation about Britney Spears' upcoming album get him to change his mind about P? It wouldn't. Consider an improvement in his attentiveness. The only thing that would do would be to cause ROBOTRON Zero to realize that Jones thought he was drinking too much at the party. But that isn't relevant to the causal origins of his beliefs or to the truth-value of P. Why would increasing ROBOTRON Zero's attentiveness, then, cause him to elicit different intuitions or to systematize differently and change his mind about P? It wouldn't. Consider an improvement in ROBOTRON Zero's intelligence. There are no unsolved puzzles whose resolution is relevant to the causal origins of his beliefs or to the truth-value of P. There are no inconsistencies in his beliefs about P or logical fallacies he committed doing his earlier systematizing. Why would increasing ROBOTRON Zero's intelligence, then, cause him to elicit different intuitions or to systematize differently and change his mind about P? Why would becoming smarter, just by itself and without any inconsistencies to discover or errors in reasoning to correct, change the intuitions he elicits or the way he systematizes them? It wouldn't. In addition, ROBOTRON Zero's conceptual repertory can't be expanded. It is already as large as it can be. He has thought of all the possible questions relevant to P that anyone could think about. He has firm intuitions about the answers to each of those questions. Thus, ROBOTRON Zero has settled on P with a priori stability. And, yet, P is false. So it is not the case that P is true if and only if it is possible for ROBOTRON Zero to on P with a priori stability. So ROBOTRON Zero does not determinately understand P.

ROBOTRON Zero's story implies that intuition is not a basic source of evidence. To see this, consider, again, the first account of Strong Modal Tie suggested by the text. According to Bealer, a candidate source of evidence has a strong modal tie to the truth if and only if, necessarily, if a subject that is in good cognitive conditions forms beliefs on the basis of a systemization of the deliverances of that source of evidence, then most of those beliefs are true.

Now, consider the way in which ROBOTRON Zero understands P. He is in very good cognitive conditions. Better than any actual human's cognitive conditions by far. Almost as good as cognitive conditions could possibly be. He has a large conceptual repertory. He has elicited intuitions about P and processed theoretically the deliverances of those intuitions. However, the resulting systemization provides an incorrect judgment about the truth of P. Of course, this first account of Strong Modal Tie to the Truth only requires that most of the beliefs about most of the TPIs that ROBOTRON Zero ends up with are true. But the example so far only holds that ROBOTRON Zero's belief about one TPI is false.

Consider, therefore, an expansion of ROBOTRON Zero's story: Everything is just as before. But the process is repeated for numerous other TPIs besides P. So there are other moral TPIs P_1 and P_2 and... and P_n. ROBOTRON Zero believes all and only these moral TPIs. The way he arrived at beliefs in P_1 through P_n is by the same method he arrived at his belief in P: Jones gave him intuitions based on the deliverances of the RTA. Those intuitions are false. ROBOTRON Zero systematizes those false intuitions about P_1 and he arrives at the wrong judgment about the truth-value of P_1. This happened in the same way as in the case of P. It happened once. There is no reason it can't happen a second time in the case of P_1. The same thing happens for P_2 and... and P_n. It happened an *i*th time. There is no reason it can happen an *i*+1th time.

Now, consider each of P, P_1, ..., P_n. In the original case, we already established that improving ROBOTRON Zero's cognitive conditions won't get him to change his mind about P. The same reasoning applies to P_1 through P_n. Consider P_1. Improving ROBOTRON Zero's memory won't get him to change his mind. Because all that would do is get him to remember a fishing trip that had nothing to do with P_1. Improving his attentiveness won't get him to change his mind either. All that would do is make him realize Jones thought he was drinking too much. Improving his intelligence won't get him to change his mind. After all, there are no inconsistencies or errors of reasoning to uncover. So no matter what improvements you make to ROBOTRON Zero's cognitive conditions, he will still believe P_1. The same goes for P_2 through P_n. Thus, all of the moral TPIs ROBOTRON Zero believes are false.

5.4.2 Intuition Does not Have a Strong Modal Tie to the Truth

Of course, there is also a second account of strong modal tie to the truth suggested by the text. Recall that according to the second account a candidate source of evidence, E, satisfies (SMTT) if and only if, necessarily, any Talkative Group that systematizes E is such that most of the beliefs of most of the members of that Talkative Group that are formed on the basis of E are true.

Now, (SMTT) only requires that most of the beliefs about most of the TPIs held by most members of a community of subjects repeating Bealer's process over a large amount of time are true. But the example so far only holds that most of ROBOTRON Zero's beliefs about most property identities are false.

Consider, therefore, a further elaboration of ROBOTRON Zero's case: Suppose that instead of just building ROBOTRON Zero, Jones builds billions of other robots—ROBOTRON 1, ROBOTRON 2, ... and ROBOTRON M. Suppose that he built each of them in the same way that he built ROBOTRON Zero. He ran the RTA. It produced a bunch of false judgments about P, P_1, … P_n. Just as he did in the case of ROBTRON Zero, Smith programmed ROBOTRON 1 through ROBOTRON M to have intuitions that match the judgments of the RTA relevant to each of those moral property identities.

Now consider ROBOTRON 1. He elicits intuitions about P. Most of those intuitions are false. He systematizes those intuitions and arrives at the incorrect judgment that P is true. It happened in the case of ROBOTRON Zero. It is possible for it to happen a second time. Suppose that ROBOTRON 1 talks to ROBOTRON Zero. Give them as much time to talk as you want. This won't get either robot to change his mind. Both robots agree that P is true. Both systematized the same intuitions about P and came to the same judgment. How could talking to each other, then, get one of them to change their mind about P? It wouldn't.

The same thing happens for P_1 through P_n. It happened with P. It is possible for it to happen again with P1. Similarly for each of P_2 through P_n. It happened an *i*th time. It can happen an *i*+1th time.

What is true of ROBOTRON 1 is true of every other member of the ROBOTRON community. So consider ROBOTRON *i*. He elicits mostly false intuitions about P. He systematizes them and arrives at the wrong judgment about P. Suppose that ROBOTRON Zero and ROBOTRON 1 and… and ROBOTRON *i* all get together and talk about P. Give them as much time to talk as you want. This won't get any of them to change their minds. All of those robots agree that P is true. All start with the same false intuitions and arrive at their judgment about P by systematizing them. How could talking to each other, just by itself, get any of the robots to change his mind about P. It wouldn't. And, again, the same reasoning holds for P_1 through P_n. So what is true of ROBOTRON Zero and ROBOTRON 1 is true of every one of the billions of other robots as well.

This ROBOTRON community is a Talkative Group. Each of them is in good cognitive conditions. No matter how much they talk to each other or how much time you give them, they are all going to believe that P and P_1 and… and P_n are true. And yet P and P_1 and… and P_n are all false. So they do not satisfy (SMTT). Thus, by Bealer's standards, intuition is not a basic source of evidence.

5.4.3 Is ROBOTRON Zero Impossible?

In response to this objection one might try denying that the ROBOTRON Zero case, or one of its elaborations, is metaphysically possible. In particular, one might deny my claim that ROBOTRON Zero wouldn't change his mind about P, or that he wouldn't change his mind about most of P, P_1, …, P_n, or that most of the ROBOTRON community wouldn't change their minds about most of P, P_1, …, P_n.

One might insist that, contrary to what I say, eventually ROBOTRON Zero would change his mind about P. Maybe this would happen after his memory becomes perfect. He remembers the fishing trip and the conversation about Britney Spears' upcoming album. And suddenly P seems false to him. Or maybe it would happen after his attentiveness is enhanced. He picks up on Smith's subtle cues, feels embarrassed about drinking too much, and then he systematizes his intuitions differently and arrives at the judgment that P is false. Or perhaps after his intelligence is increased enough, it will just seem to him that P is false. In general, it is metaphysically necessary that if you keep improving a subject's cognitive conditions, then that subject will eventually have intuitions that yield mostly true judgments. The case of ROBOTRON Zero, as I have described it, is impossible.

Or perhaps one might insist that the story I tell could be true of one or two or three robots. But if you keep making ROBOTRONs, eventually you will get to one that has different intuitions than the others. That ROBOTRON switches intuitions, talks to the rest of the ROBOTRON community, and as a result of that conversation the ROBOTRON community is led down a path of epistemological redemption—trading in their false beliefs about the relevant TPIs for true beliefs about those TPIs.

The problem with this response is that Bealer promises to give us a theory that explains why there is a strong modal tie between intuition, phenomenal experience, and the truth. His complaint about Empiricism and Rationalism is that they simply took for granted that phenomenal experience and intuition are tied to the truth without explaining why. He (1999, p. 36-7) says, for example, this:

> A shortcoming of traditional empiricism was that it offered no explanation of why phenomenal experience is a basic source of evidence; this was just an unexplained dogma. By the same token, traditional rationalists (and also moderate empiricists who, like Hume, accepted intuition as a basic source of evidence) did not successfully explain why intuition is a basic source of evidence. Modal reliabilism provides a natural explanation filling in these two gaps. The explanation is in terms of the indicated modal tie between these sources and the truth. But why should there be such a tie to the truth? Neither traditional empiricism nor traditional rationalism provided a satisfactory explanation. The theory of concept possession promises to fill in this remaining gap.

And later he (1999, p. 47) says:

> In the course of our discussion of the evidential force of intuitions, we noted a shortcoming in traditional empiricism and traditional rationalism, namely, that neither successfully explains why intuition and phenomenal experience should be basic sources of evidence. Modal reliabilism filled this explanatory gap: the explanation is that these two sources of evidence have the right sort of modal tie to the truth. We saw, moreover, that neither traditional empiricism nor traditional rationalism successfully explains why there should be such a tie between these basic sources and the truth. The analysis of determinate concept possession fills this gap.

Now, if Bealer resorts to denying that cases like ROBOTRON Zero's or one of its expansions is metaphysically possible, then his theory suffers from the same problem he identifies in traditional empiricism and rationalism. Rather than explaining why there is a strong modal tie between intuition and truth, Bealer simply assumes that there is such a tie and hides that assumption behind a complex series of definitions. It is just a brute metaphysical fact that if the members of a Talkative Group are smart enough and attentive enough and mnemonically gifted enough, most of their members will have intuitions that are roughly and for the most part true. If members of such a group are reasonably smart and attentive and mnemonically gifted, but they believe a false TPI, then most of those members have not settled on that TPI with a priori stability. If they are made smarter or more attentive or more mnemonically gifted, it is metaphysically necessary that most of them will eventually receive new intuitions and change their minds about that TPI.

This is not an explanation. It is an undefended assumption. Furthermore, this assumption, all by itself, without any of the complex definitions Bealer introduces is enough to get the result that Bealer wants. So if Bealer is entitled to this assumption, then why aren't old fashioned Empiricists and Rationalists entitled to it as well? And if those old fashioned theorists aren't entitled to the relevant assumption, then why is Bealer?

BIBLIOGRAPHY

Adams, Robert (1979). 'Divine Command Metaethics Modified Again'. *The Journal of Religious Ethics*, 7, 66-79.

Adams, Robert (1999). *Finite and Infinite Goods: A Framework for Ethics*. Oxford University Press.

Adams, Robert (2003). 'Anti-Consequentialism and the Transcendence of the Good'. *Philosophy and Phenomenological Research*, 67, 114-32.

Adams, Robert (2006). *A Theory of Virtue: Excellence in Being for the Good.* Oxford University Press.

Bealer, George (1987). 'The Philosophical Limits of Scientific Essentialism'. *Philosophical Perspectives,* 1, 289-365.

Bealer, George (1996a). 'A Priori Knowledge and the Scope of Philosophy'. *Philosophical Studies*, 81, 121-142.

Bealer, George (1996b). 'A Priori Knowledge: Replies to Lycan and Sosa'. *Philosophical Studies*, 81, 163-174.

Bealer, George (1998a). 'A Theory of Concepts and Concept Possession'. *Philosophical Issues*, 9, 261-301.

Bealer, George (1998b). 'Concept Possession'. *Philosophical Issues*, 9, 331-338.

Bealer, George (1999). 'A Theory of the A Priori'. *Philosophical Perspectives,* 13, 29-55.

Bealer, George (2000). 'A Theory of the A Priori'. *Pacific Philosophical*

Quarterly, 81, 1-30.

Boyd, Richard (1988). 'How to Be a Moral Realist'. in Sayre-McCord, *Essays on Moral Realism*, 181-228.

Boyd, Richard (2003a). 'Finite Beings, Finite Goods: The Semantics, Metaphysics and Ethics of Naturalist Consequentialism, Part I'. *Philosophy and Phenomenological Research*, 66, 505-53.

Bradley, Ben (2002). 'Is Intrinsic Value Conditional?'. *Philosophical Studies*, 107, 23-43.

Brink, David (2001). 'Realism, Naturalism, and Moral Semantics'. *Social Philosophy and Policy*, 18, 154–176.

Brown, Jessica (1998). ‘Natural Kind Terms and Recognitional Capacities’. *Mind*, 107, 275-303.

Copp, David (2000). ‘Milk, Honey, and the Good Life on Moral Twin Earth’. *Synthese*, 124, 113–137.

Devitt, Michael and Sterelny, Kim (1999). *Language and Reality*. The MIT Press.

Evans, Gareth (1973). ‘The Causal Theory of Names’. *Proceedings of the Aristotelian Society, Supplementary Volumes*, 47, 187-208.

Feldman, Fred (1983). *Doing the Best We Can: An Essay in Informal Deontic Logic*, Riedel.

Feldman, Fred (1995). ‘Adjusting Utility for Justice’. *Philosophy and Phenomenological Research*, 55, 567-585.

Feldman, Fred (1997). *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*. Cambridge University Press.

Feldman, Fred (1998). ‘Hyperventilating About Intrinsic Value’. *The Journal of Ethics*, 2, 339-354.

Feldman, Fred (2002). ‘The Good Life: A Defense of Attitudinal Hedonism’. *Philosophy and Phenomenological Research*, 65, 604-628.

Feldman, Fred (2004). *Pleasure and the Good Life: Concerning the Nature, Varieties, and Plausibility of Hedonism*. Oxford University Press.

Feldman, Fred (2005). ‘The Open Question Argument: What it Isn’t; and What it Is’. *Philosophical Issues*, 15, 22-43.

Harris, John (1975). ‘The Survival Lottery’. *Philosophy,* 50, 81-87.

Horgan, Terence and Timmons, Mark (1990–91). ‘New Wave Moral Realism Meets Moral Twin Earth’. *Journal of Philosophical Research*, 16, 447–465.

Horgan, Terence and Timmons, Mark (1992a). ‘Troubles on Moral Twin Earth: Moral Queerness Revived’. *Synthese*, 92, 221–260.

Horgan, Terence, and Timmons, Mark (1992b). ‘Troubles for New Wave Moral Semantics: The Open Question Argument Revived’. *Philosophical Papers*, 21, 153–175.

Horgan, Terence and Timmons, Mark (2000). ‘Copping out on Moral Twin Earth’. *Synthese*, 124, 139–152.

Huemer, Michael (2005). *Ethical Intuitionism*. Palgrave Macmillan.

Jenkins, Carrie (2008). *Grounding Concepts: An Empirical Basis for Arithmetic Knowledge*. Oxford University Press.

Kagan, Shelly (1998). ‘Rethinking Intrinsic Value’. *The Journal of Ethics*, 2, 277-297.

Kagan, Shelly (2009a). ‘The Grasshopper, Aristotle, Bob Adams, and Me’. in Newlands and Jorgensen, *Metaphysics and the Good: Themes from the Philosophy of Robert Merrihew Adams*. Oxford University Press.

Kagan, Shelly (2009b). ‘Well-Being as Enjoying the Good’. *Philosophical Perspectives*, 23, 253-272.

Kornblith, Hilary (2006). ‘Intuitions and the Ambitions of Philosophy’. in Hetherington, *Epistemology Futures*. Oxford University Press.

Kripke, Saul (1980). *Naming and Necessity*. Harvard University Press.

Laurence, Stephen and Margolis, Eric (2003). ‘Concepts and Conceptual Analysis’. *Philosophy and Phenomenological Research*, 67, 253-82.

Levy, Neil (forthcoming). ‘Moore on Twin Earth’. *Erkenntnis,* 1-10.

Lewis, David (1973). ‘Counterfactuals and Comparative Possibility’. *Journal of Philosophical Logic*, 2, 418-46.

Lewis, David (1979). ‘Counterfactual Dependence and Time’s Arrow’. *Nous*, 13, 455-76.

Lewis, David (1986). *Philosophical Papers: Volume II.* Oxford University Press.

Merli, David (2002). ‘Return to Moral Twin Earth’. *Canadian Journal of Philosophy*, 32, 207–240.

Putnam, Hilary (1975). *Mind, Language, and Reality: Philosophical Papers Volume 2*. Cambridge University Press.

Rea, Michael. (2006). ‘Naturalism and Moral Realism’. in Thomas M. Crisp, Matthew Davidson, David Vander Laan eds. *Knowledge and Reality: Essays in Honor of Alvin Plantinga*. Dordrecht: Kluwer.

Rubin, Michael (2008a). ‘Sound Intuitions on Moral Twin Earth’. *Philosophical Studies,* 139, 307–327.

Rubin, Michael (2008b). ‘Is Goodness a Homeostatic Property Cluster?’. *Ethics*, 118, 496-528.

Sayre-McCord, Geoff. (2009). ‘Moral Realism’. *The Stanford Encyclopedia of Philosophy.*

Sarch, Alexander. (2010). ‘Bealer and the Autonomy of Philosophy’. *Synthese*, 174, 451-474.

Schechter, Joshua (2010). ‘The Reliability Challenge and the Epistemology of Logic’. *Philosophical Perspectives*, 24, 437-464.

Sider, Theodore (1991). ‘Might Theory X Be a Theory of Diminishing Marginal Value?’. *Analysis*, 51, 265-271.

Sider, Theodore (1993). ‘Asymmetry and Self-Sacrifice’. *Philosophical Studies*, 70, 117-132.

Singer, Peter (1977). Utility and the Survival Lottery. *Philosophy*, 52, 218-222.

Soames, Scott (2002). *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford University Press.

Stanford, Kyle and Kitcher, Philip (2000). ‘Refining the Causal Theory of Reference for Natural Kind Terms’. *Philosophical Studies*, 97, 99-129.