Goodness and Moral Twin Earth

Christopher Freiman
College of William and Mary
Department of Philosophy

Moral Twin Earth, a thought experiment advanced by Terry Horgan and Mark Timmons, threatens the very possibility of naturalist moral realism. Naturalist moral realists claim that moral facts are entailed by natural facts. The thought experiment purportedly shows that any naturalist realist moral semantics that yields sufficiently determinate referents will imply that moral facts differ across worlds in which the relevant natural facts do not differ. If successful, Moral Twin Earth defeats naturalist moral realists' entailment claim.

Moral Twin Earth causes problems for a variety of theories of naturalist moral realism including new wave moral realism,[1] analytical moral functionalism,[2] and dispositional theories of value,[3] among others. Indeed, Horgan and Timmons argue that it undercuts "virtually any metaphysically naturalist version of moral realism."[4]

A satisfactory reply to Moral Twin Earth may be found in an idea that some have been urging for the better part of a century. I will sketch an Aristotelian moral semantics that is unique in construing terms like 'right' and 'good' exclusively as attributive adjectives that denote relational properties. On this view, moral goodness is a relational property predicated of those human beings that satisfy kind-relative criteria of goodness: i.e., morally good human beings are good *qua* human being. The reference of moral terms is therefore fixed by the natural properties of human beings. This account ensures that moral facts cannot differ

---

[1] Horgan and Timmons, 1991; Horgan and Timmons, 1992a; Horgan and Timmons 1992b.
[2] Horgan and Timmons, 2009.
[3] Holland, 2001.
[4] Horgan and Timmons, 2009, p. 221.

across worlds in which the relevant natural facts do not differ and thus defeats Moral Twin Earth.

Before proceeding, it is worth clarifying that the aim of this paper is not to vindicate Aristotelian ethical naturalism generally. Rather, I wish to show that Aristotelian ethical naturalism, a well-established form of naturalist moral realism, is not vulnerable to the specific challenge posed by the Moral Twin Earth thought experiment—a challenge which purports to be decisive against *all* forms of naturalist moral realism. However, I leave open the possibility that Aristotelian ethical naturalism fails for other reasons.

First I situate the Moral Twin Earth thought experiment in historical context by delineating G.E. Moore's open question argument. Moral Twin Earth is a "revived" version of Moore's argument against the viability of analytically true naturalistic definitions of moral terms; a satisfactory reply to Moral Twin Earth is substantively similar to the Aristotelian reply to the open question argument which originated with Peter Geach. I then detail Horgan and Timmons's updating of the open question argument in their treatment of "Cornell" realism [1]. I argue that many naturalist moral realists are vulnerable to Moral Twin Earth because they accept that moral terms such as 'right' and 'good' are predicative adjectives that denote intrinsic properties rather than attributive adjectives that denote relational properties [2]. I proceed to sketch an Aristotelian role semantic account of moral terms [3]. I then apply this account to the problem of Moral Twin Earth and conclude that it satisfies the desiderata that a reply to Moral Twin Earth must satisfy, as specified by Horgan and Timmons themselves [4]. I close by noting that there remain significant unresolved challenges to neo-Aristotelian ethical naturalism [5].

[1]

The open question argument aims to show that moral properties cannot be identical with natural properties.[5] This conclusion is entailed by the premises: (1) if moral terms lack analytically true naturalistic definitions, then moral properties are not identical with any natural properties, and (2) moral terms lack analytically true naturalistic definitions.

Moore argues for the second premise thusly: if 'goodness' is to be analyzed as the natural property of, say, pleasantness, we should be able to substitute the term 'pleasantness' for any occurrence of the term 'goodness' without a loss of meaning. The term 'goodness' cannot mean pleasantness unless the question "Is goodness pleasantness?" is closed. That is, any competent English speaker would answer "yes", just as she would answer "yes" to the question, "Is a bachelor an unmarried man?"

But competent English speakers can always ask "*Is* pleasantness goodness?" The question is not closed. Thus, the term 'good' does not have an analytically true naturalistic definition. If we also accept Moore's first premise—if moral terms lack analytically true naturalistic definitions, then moral properties are not identical with any natural properties— we must conclude that moral properties are not identical with any natural properties.

The "new wave" or Cornell moral realists resist Moore's conclusion by rejecting his first premise.[6] They argue that moral terms can have *synthetically* true naturalistic definitions. Richard Boyd, for instance, appeals to the causal theories of reference developed by Saul Kripke[7] and Hilary Putnam.[8] These theorists deny that the reference of certain terms (e.g. names and natural kind terms) is determined by a description with which the terms are associated by competent speakers. For instance, the referent of 'water' is not determined by a description like "the clear liquid that fills the oceans."

---

[5] Moore, 1988.
[6] See, for example, Boyd, 1988; Sturgeon, 1985; Brink, 1989.
[7] Kripke, 1980.
[8] Putnam, 1975.

Putnam posits Twin Earth to support this claim. Twin Earth is identical to Earth with one exception. On Twin Earth

> the liquid called 'water' is not H_2O but a different liquid whose chemical formula is very long and complicated. I shall abbreviate this chemical formula simply as XYZ. I shall suppose that XYZ is indistinguishable from water at normal temperatures and pressures. In particular, it tastes like water and quenches thirst like water. Also, I shall suppose that the oceans and lakes and seas of Twin Earth contain XYZ and not water, that it rains XYZ on Twin Earth and not water, etc.[9]

Putnam claims that the referent of 'water' differs on Earth and Twin Earth. On Earth, 'water' refers to H_2O. On Twin Earth, 'water' refers to XYZ.[10]

The referent of the term 'water' is determined by a certain kind of causal relation between the referent and the use of the term. Boyd writes,

> *Roughly*, and for nondegenerate cases, a term t refers to a kind (property, relation, etc.) k just in case there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term t will be approximately true of k (excuse the blurring of the use-mention distinction). Such mechanisms will typically include the existence of procedures which are approximately accurate for recognizing members or instances of k (at least for easy cases) and which relevantly govern the use of t, the social transmission of certain relevantly approximately true beliefs regarding k, formulated as claims about t (again excuse the slight to the use-mention distinction), a pattern of deference to experts on k with respect to the use of t, etc. . . . When relations of this sort obtain, we may think of the properties of k as regulating the use of t (via such causal relations) and we may think of what is said using t providing us with a socially coordinated *epistemic access* to k: t refers to k (in nondegenerate cases) just in case the socially coordinated use of t provides significant epistemic access to k, and not to other kinds (properties, etc.).[11]

The term 'water,' for instance, refers to the property that is causally related to our use of the term 'water' in the way Boyd describes. When Boyd's conditions obtain, we can follow Horgan and Timmons and say that H_2O "causally regulates" the use of the term 'water.'

---

[9] Ibid., p. 223.

[10] It's possible that competent speakers on Earth associate the description "being composed of H_2O" with the term "water." Similarly, competent speakers on Twin Earth may associate the description "being composed of XYZ" with the term "water." Thus, Putnam asks us to imagine both Earth and Twin Earth as they were in 1750, before the development of modern chemistry. In 1750, Earthlings and Twin Earthlings would not associate "water" with descriptions of its chemical composition. Rather they will associate similar, if not identical, descriptions with the term "water," e.g., "the clear liquid that fills the oceans." Putnam claims that even in 1750, the term "water" referred to H_2O on Earth and XYZ on Twin Earth. Ibid., p. 224.

[11] Boyd, 1988, p. 195. Boyd has since revised his semantics. For his latest account, see Boyd, 2003.

Boyd's semantics explain how competent speakers can doubt the identity of two properties that are actually identical. The proposition that "water is H_2O" is discoverable only a posteriori. Knowledge that water is H_2O cannot be acquired through conceptual analysis of the terms 'water' or 'H_2O,' or through a priori intuition.

We need empirical inquiry, not merely mastery of language, to determine whether 'water' refers to H_2O. Natural science informs us that the property of being H_2O causally regulates our use of the term 'water.' Competent speakers (and even experts, depending on the state of scientific theorizing) might not know the property that causally regulates the use of a given term. Thus, open questions about identical properties are possible.

Boyd uses the causal theory of reference to defend the claim that basic moral terms can have synthetically true naturalistic definitions. He analogizes moral terms to natural kind terms. 'Good' refers to a natural property in just the way that 'water' refers to the natural property of H_2O. The term 'good' refers to the natural property (or perhaps the cluster of natural properties) that causally regulates our use of the term 'good.'

Competent speakers can doubt the identity of goodness and a given natural property even when the two properties are actually identical, just as they can doubt the identity of water and H_2O. Since discovering the property that causally regulates the uses of our moral terms can require extensive empirical inquiry, competent speakers (and even experts) can be unaware of the identity. Thus, the new wave moral realist concludes that open questions about the identity of a moral property and a natural property are possible even when the properties are identical.

Horgan and Timmons argue that a revised version of the open question argument undermines new wave moral realism. First, they stipulate that on Earth, human uses of basic moral terms like 'good' and 'right' are causally regulated by certain natural properties, namely

consequentialist properties whose "essence is captured by some specific consequentialist normative theory; call this theory T^c."[12]

Next, Horgan and Timmons posit Moral Twin Earth. Moral Twin Earth is virtually identical to Earth. Twin Earthlings use moral terms that are orthographically identical to Earthling moral terms and in the same way as Earthlings. The Twin Earthlings use the terms 'right and wrong,' 'good and bad' to judge actions, traits, and social practices, as well as to reason about well-being. Twin Earthlings are typically disposed to act in conformity with their judgments of what is 'good' and 'right' and react favorably to those actions, traits, and practices they judge 'good' and 'right' and disfavorably to those actions, traits, and practices they judge 'bad' and 'wrong.'[13]

But Earth and Moral Twin Earth differ in one important respect: on Moral Twin Earth, the use of basic moral terms is causally regulated by deontological properties whose essence is captured by some specific deontological normative theory "T^d." The difference, Horgan and Timmons stipulate, is due to species-wide differences in psychological temperament (for instance, Twin Earthlings experience guilt more intensely, and sympathy less intensely, than Earthlings).[14]

Horgan and Timmons propose two alternative characterizations of the uses of moral terms on Earth and Moral Twin Earth. Either moral and twin moral terms differ in meaning or they do not.[15] If the meanings do not differ, new wave moral semantics is false. According to new wave moral semantics, moral and twin moral terms *must* differ in meaning because different properties causally regulate the Earthling and Twin Earthling use of moral terms.

---

[12] Horgan and Timmons, 1991, p. 458.
[13] Ibid., p. 459.
[14] Ibid.
[15] Ibid., p. 460.

But intuitively, moral and twin moral terms do not differ in meaning. Timmons asks us to consider how discussion between Earthlings and Twin Earthlings would proceed. Suppose the Earthlings and Twin Earthlings discover that different properties causally regulate their uses of moral terms. If new wave moral semantics were correct, both sides ought to regard intergroup debate about goodness as inappropriate or foolish; they would construe their dispute as purely semantic. But surely both sides would claim that intergroup debate is *not* inappropriate, which indicates that the dispute concerns moral belief rather than meaning.[16] The disputants will think that they are discussing the same property and simply ascribe false beliefs about the property to each other.

Thus, the key intuition here is that the Earthlings and Twin Earthlings are engaged in a genuine moral disagreement. That is, their dispute is a substantive one about the correct account of moral properties. The Earthling contends that consequentialism is correct and that the Twin Earthlings, who adhere to some version of deontology, are mistaken to think otherwise. The Twin Earthlings will contend that the consequentialist Earthling is, in fact, the mistaken one. The critical point is that the parties are not simply talking past each other.

To clarify further why we ought to regard the dispute between Earthlings and Twin Earthlings as a genuine moral disagreement, consider the following contrast between Putnam's Twin Earth and Moral Twin Earth. An Earthling on Putnam's Twin Earth will not insist that H_2O is the *true* composition of water once she learns that XYZ causally regulates the Twin Earthling use of the term 'water.' This revelation will prompt her to withdraw her objection to the Twin Earthlings. In contrast, a consequentialist Earthling on Moral Twin Earth will insist that we *really should* maximize utility (for example) even upon learning that deontological properties causally regulate the use of twin moral terms. The Earthling will not

---

[16] Timmons, 1999, p. 62ff.

withdraw her objection to the Twin Earthlings; instead she will maintain that Twin Earthlings are making a substantive error about morality.

The dispute between the Earthling and Twin Earthling on Putnam's Twin Earth will dissolve because both disputants will accept that the term 'water' simply refers to different properties on Earth and Twin Earth. But the dispute between the Earthling and the Twin Earthling on Moral Twin Earth will *not* dissolve because neither disputant will accept that the term 'good' refers to different properties on Earth and Twin Earth. The disputants will claim that the term refers to the *same* property—the property of goodness—and that the other party has false beliefs about that property.

New wave moral semantics entails that the meaning of moral and twin moral terms differ. Intuitively, Moral Twin Earth shows that their meanings do not differ. If we share this intuition, we will conclude that new wave moral semantics is false. Here we see the open question argument revived: even though a given natural property may causally regulate our use of the term 'good,' we can question whether that property *really is* good.

Horgan and Timmons conceive of Moral Twin Earth as a generic argument that forces naturalists of all stripes to accept either indeterminacy or relativism.[17] The argument will always take the following form: first, Horgan and Timmons grant the realist's account of the relation between basic moral terms like 'good' and 'right' and the natural property (or cluster of natural properties) to which they refer. The account is guilty of indeterminacy if it allows that natural properties specified by incompatible moral theories—e.g., deontological *and* consequentialist *and* virtue ethical properties—are eligible referents for moral terms like 'good' and 'right.'

---

[17] Horgan and Timmons, 2000.

Any account of moral semantics determinate enough to restrict the reference of moral terms to a suitably circumscribed set of natural properties (e.g., consequentialist properties) will be guilty of relativism. To illustrate, Horgan and Timmons posit Moral Twin Earth, a planet virtually identical to Earth in terms of natural facts on which human beings use putatively moral terms (terms that are used to regulate action, reason about well-being, and so forth) to refer—according to the realist's own account of the reference of moral terms—to properties outside of the eligible set (e.g., deontological properties).

The realist has two options. First, she might claim that, in spite of appearances, the terms 'good' and 'right' *mean* something different on Earth and Moral Twin Earth. However, as evidenced by the case of new wave moral semantics, this reply is counterintuitive. Alternatively, the realist might claim that moral terms have the same meaning on Earth and Moral Twin Earth, but different referents. This view implies that the same moral judgment can be true for Earthlings and false for Twin Earthlings; Twin Earthlings are using genuinely moral terms to refer to properties that are good or right *for them*. This reply is an unacceptable option for naturalist moral realists, who claim that the moral facts are entailed by the natural facts. Thus, a viable naturalist moral realism must (1) secure significant determinacy of moral reference and (2) avoid relativism.

[2]

In his treatment of the open question argument, Peter Geach argues that 'good' is an attributive adjective; i.e., the criteria for the term's proper application are fixed by a corresponding noun phrase.[18] For instance, 'good' in 'Spinoza grinds good lenses' applies correctly only in virtue of what it is to be a lens. Geach claims that this grammatical feature

---

[18] Geach, 1967.

of 'good' does not change in moral contexts, only the corresponding noun phrase (i.e., 'human being') does.

Geach writes, "I shall say that in a phrase 'an AB' ('A' being an adjective and 'B' being a noun) 'A' is a (logically) predicative adjective if the predication 'is an A B' splits up logically into a pair of predications 'is a B' and 'is A'; otherwise I shall say that 'A' is a (logically) attributive adjective."[19] 'Gold' is a predicative adjective. "Blackbeard's booty is a gold doubloon" analyzes into "Blackbeard's booty is gold" and "Blackbeard's booty is a doubloon." All gold things share an intrinsic property in virtue of which they are gold—they are composed of material with an atomic number of 79.

'Big' is an attributive adjective. "Antigone is a big ant" does not analyze into "Antigone is big" and "Antigone is an ant." A big ant and a big elephant do not share an intrinsic property in virtue of which they are both big. Rather they share a relational property: they are both big relative to a kind-specific reference class.

For Moore, 'good' is a predicative adjective like 'gold.'[20] For Geach, 'good' is an attributive adjective—we cannot know the property to which 'good' refers in the expression 'good α' until we know what kind of thing α is. The reference of 'good' is indeterminate unless some criterion of evaluation is in place.[21] The criterion of evaluation is supplied by the noun phrase 'good' modifies. Something's goodness is determined by kind-relative criteria,

---

[19] Ibid., p. 64.

[20] Moore writes, "Ethics is undoubtedly concerned with the question what good conduct is; but, being concerned with this, it obviously does not start at the beginning, unless it is prepared to tell us what is good as well as what is conduct. For 'good conduct' is a complex notion: all conduct is not good; for some is certainly bad and some may be indifferent. And on the other hand, other things, besides conduct, may be good; and if they are so, then, 'good' denotes some property, that is common to them and conduct; and if we examine good conduct alone of all good things, then we shall be in danger of mistaking for this property, some property which is not shared by those other things." Moore, 1988, p.2.

[21] A similar analysis might hold for "right." No property constitutes the right move in a chess match, the right program for installing an operating system, or the right bit for a given drill. See Thomson, 2006.

just as its bigness is determined by kind-relative criteria. Judith Jarvis Thomson argues that 'good' does not typically function as a predicative adjective in ordinary linguistic practice, noting that philosophers like Geach and G.H. von Wright[22] "gave the excellent advice that we should *look and see* how the word 'good' is actually used. They showed, conclusively, that it does not function in the way in which adjectives like 'visible' and 'happy' do."[23]

The Aristotelian might argue that we do not agree on *the* criteria for the application of the term 'good' because there are no such criteria. No account of moral semantics will yield a determinate referent for the predicate 'good' because 'good' only has determinate criteria of application when it is construed as an attributive adjective and prefaces a substantive. On this view, new wave moral realists fail to resist Moore's conclusion that no natural property is identical with the property of goodness because they fail to deny Moore's crucial premise: there is an intrinsic property of goodness in virtue of which all good things are good just as there is property of being H_2O in virtue of which all water is water.

Here is an analogy. The question, "Is having been painted by Vermeer identical to being genuine?" seems open. Competent speakers can doubt whether having been painted by Vermeer is identical to being genuine. One good reason for doubting this identity relation is that having been painted by Vermeer *isn't* identical to being genuine. 'Genuine' is an attributive adjective; therefore, it is not the case that all genuine things share an intrinsic property in virtue of which they are genuine. Rather they relate to things of their kind in a certain way such that they are genuine. A gold doubloon, for instance, can be genuine without having been painted by Vermeer. A genuine gold doubloon is genuine in virtue of being composed of atoms having an atomic number of 79. Nothing—that is, no single

[22] Wright, 1963.
[23] Thomson, 1997, p. 275. Thomson's account differs from Geach's in certain respects, but I omit discussion of such details because they are orthogonal to my purpose here.

intrinsic property—is identical to being genuine. Any such property one suggests as being identical to genuineness will succumb to the open question argument. The search for the single intrinsic natural (or non-natural) property that is identical with goodness is analogous to a search for the single intrinsic natural property that is identical with genuineness. If 'good,' like 'genuine,' is an attributive adjective, then any such property (or cluster of such properties) we propose as being identical with goodness will succumb too.

[3]

Geach's reply to Moore is unmistakably Wittgensteinian: puzzlement about goodness arises when we stray from noting the ways 'good' is, as a matter of fact, used in ordinary linguistic practice. Wittgenstein was a forerunner of contemporary role accounts of meaning and Geach's analysis of the open question argument suggests that we can profitably construct an Aristotelian account of moral semantics along these lines.

Role semantic accounts of moral terms are not new. Michael Smith[24] and Frank Jackson,[25] for instance, present theories whereby the meaning of a moral term like 'good' is determined by its role in a system of platitudes that are constitutive of our ordinary uses of the term (e.g., "if an action is good, it is worthy of praise"). The term 'good' refers to whatever natural properties happen to fill the role of the term 'good' in the system of platitudes. Horgan and Timmons have already deployed Moral Twin Earth to counter this view and I will not articulate it further.[26]

The Aristotelian objection to this type of view can be inferred from earlier remarks: it treats moral evaluation as if it were *sui generis*, distinct from other forms of evaluation. Smith

---

[24] Smith, 1994.
[25] Jackson, 1998.
[26] Horgan and Timmons, 2009.

and Jackson analyze the functional role of moral terms without considering their place in a comprehensive evaluative vocabulary.

According to the moral semantics implicit in the work of philosophers like Geach and Thomson, evaluative terms and concepts *broadly construed* are functionally definable by a set of platitudes. The meaning of the term 'good' is determined by its role in a series of platitudinous generalizations such as "something is a good α if it is an exemplary token of the type α," "good α's function efficaciously," and "when choosing α's, normal agents choose good α's." Thus, the Aristotelian agrees with the functional role theorist about how the term 'good' gets its meaning, but disagrees about the relevant set of platitudes for the term.

Aristotelians consider functional evaluations paradigmatic ('good knife,' 'good swimmer,' etc.), so these are the cases I will focus on.[27] The Aristotelian view seems to hold that if 'β' is a functional concept, the *essential* role of the term 'good' in the expression 'good β' is to individuate well-functioning tokens of the type β.[28] To avoid circularity, I understand "well-functioning tokens of type β" roughly to be those that meet some β-relative threshold of functionality: (e.g.) a good knife is a knife that meets some threshold of cutting ability.

Similarly, we evaluate the goodness of plants and animals according to kind-relative functional criteria. Good organisms are those fitted with the traits needed to achieve the characteristic functions of their kind. Roughly speaking, we evaluate an organism as a good

---

[27] There are, of course, instances of the predicate 'good α' in which α's are not functional concepts, and thus, the criterion of evaluation is something other than functionality.

[28] For my purposes here, when 'α' is a functional concept, I will focus on the function of α, ignoring other semantic characteristics it might have. At least part of its semantic meaning is its functional role. There might be more to consider, e.g., part of the meaning of the term 'guitar' is given not only by its function, but also by its history. To simplify, I will simply focus on functions.

specimen of its kind by how efficaciously its parts and operations contribute to the teleonomic functions characteristic of that kind of organism.[29]

Ethical evaluations of human beings proceed similarly. Philippa Foot, for instance, claims that the meaning of 'good' does not change in moral contexts.[30] The criteria for the proper application of 'good' in 'honesty is a morally good disposition' belong to the same class as the criteria for the proper application of 'good' in 'this tree has good roots.' What changes in moral contexts is not the meaning of 'good' but the corresponding noun phrase (i.e., 'human being').

Aristotelian naturalism's signature thesis holds that evaluations of human beings are anchored by the characteristic functions of the species. Aristotelian-minded moral philosophers have supplied an expansive and varied menu of options for specifying goodness *qua* human.[31] What these theorists share is the belief that morality is grounded, in some way, in behaviors and ends characteristic of human beings.

Space precludes a detailed exposition of any particular Aristotelian account of human functioning and the account of human virtue grounded therein. I elide points of controversy (both intermural and intramural) about the proper specification of characteristic human functioning and human virtue because they are orthogonal to the specific problems generated by Moral Twin Earth. My aim here is not to defend a particular specification against all competitors or to resolve the difficulties attendant on any proposed specification of human functioning; therefore, I will leave the details of the various proposals aside and simply direct the reader to the more complete specifications and defenses provided in the work of Aristotelian moral theorists themselves. I intend only to show that Aristotelian

---

[29] For a more detailed account, see Thomson, 1997.
[30] Foot, 2003, p. 39.
[31] See, for instance, Casebeer, 2005; Foot, 2003; Hursthouse, 1999; Nussbaum, 2000.

ethical naturalism is not defeated by Moral Twin Earth—a thought experiment that allegedly defeats any form of naturalist moral realism. I want to emphasize that this does not mean that Aristotelian ethical naturalism is vindicated generally.

[4]

We are now in a position to explore the Aristotelian reply to Moral Twin Earth. The first horn of Moral Twin Earth's dilemma is indeterminacy of moral reference. Recall that Horgan and Timmons challenge the naturalist moral realist to offer an account of reference-fixing for moral terms that fixes determinate reference-relations between moral terms and specific natural properties. If a theory's semantic constraints are satisfied by deontological, consequentialist, and virtue ethical properties, it is insufficiently determinate.

The Aristotelian theory's semantic constraints seem capable of ruling out competing properties. Utilitarianism furnishes an example. Foot insists that "*utilitarianism never gets off the ground*" on the Aristotelian account of goodness.[32] Geach likewise argues that "we cannot sensibly speak of a good or bad event, a good or bad thing to happen. 'Event,' like 'thing,' is too empty a word to convey either a criterion of identity or a standard of goodness."[33] The Aristotelian account anchors evaluation to kind-relative functional criteria; states of affairs *tout court* are therefore not the kind of thing that can supply determinate criteria of application for the attributive adjectives 'good' and 'bad.' The basic Aristotelian objection to utilitarianism, in brief, is that it takes the predicative use of 'good' to be fundamental and thus is grounded in (illicit) evaluations of states of affairs. For this reason, utilitarian properties are ruled out as eligible referents for terms like 'good' and 'bad.'[34]

---

[32] Foot, 2003, p. 49, italics in the original.

[33] Geach, 1967, p. 72.

[34] One might dispute that evaluations of states of affairs *tout court* are essential to utilitarianism or consequentialism broadly construed. For the claim that states of affairs are the locus of utilitarian evaluation, see Thomson, 1994.

Rosalind Hursthouse notes that if we are using the term 'good' attributively, we must evaluate objects by the criterion appropriate to each kind of object. In the context of living organisms, the appropriate criterion is roughly its success in achieving its type's characteristic ends. Hursthouse says the idea that 'good' is an attributive adjective entails that

> although you can evaluate and choose things according to almost any criteria you like, you must select the noun or noun phrase you use to describe the thing you are calling good *advisedly*, for it determines the criteria of goodness that are appropriate. [R.M.] Hare can call a cactus a good one on the grounds that it is diseased and dying, and choose it for that reason, but what he must not do is describe it as a good *cactus*, for a cactus is a living thing. He can describe it as a good 'decorative object for my windowsill' or 'present to give my detestable mother-in-law,' but not as a good *cactus*.[35]

A cactus can be a good gift (that is, good *qua* gift) without being a good *cactus* (that is, good *qua* cactus). An object like a knife or a cactus or a human being "brings its own criteria of goodness with it."[36] We can call a cactus a good one because it is soft and doughy and makes for a good pillow. But we could not properly call such a thing a good *cactus*.

Similarly, we can call a human being a good one by any standard we like, including how well she maximizes utility. But we could not properly call such a person a good *human being*. The appropriate criterion for evaluating goodness *qua* human being is the characteristic functioning of human beings.[37]

The preceding indicates that Aristotelianism's semantic constraints cannot be satisfied by (one kind of) consequentialism. Of course, I don't take the preceding to *establish* that moral terms, on the Aristotelian view, are sufficiently determinate to disallow the natural properties suggested by rival moral theories from being eligible referents. There are, after all, many

---

[35] Hursthouse, 1999, p. 195.
[36] Ibid., p. 197.
[37] It is possible that, as a contingent matter of fact, humans who pursue characteristic human ends maximize utility. However, in this case, the fact that good humans in the Aristotelian sense are also good humans in the utilitarian sense would be coincidental. On the utilitarian view, there is no necessary connection between being a good human and functioning in characteristically human ways. Similar remarks can be made about deontology. As a contingent matter of fact, characteristic human ends might align with Kantian moral duties. However this overlap would be purely coincidental.

variants of these theories. But I have offered considerations that suggest the Aristotelian account of moral semantics is strong enough to pin down determinate referents of moral terms.

The Moral Twin Earth thought experiment enters once a theory shows that it can generate significant determinacy of moral reference. Recall Moral Twin Earth's structure. Horgan and Timmons begin by granting the realist's semantic story. Then, they posit Moral Twin Earth—a planet similar to Earth in terms of natural facts where the realist's account of moral semantics holds, but where the natural property to which moral terms are properly related differs from the natural property to which moral terms are properly related on Earth.

The realist has two choices. First, she might claim that, in spite of appearances, there is no genuine moral disagreement between Earthlings and Twin Earthlings. This response implies that the orthographically identical Earthling and Twin Earthling terms have different meanings and different referents. The Earthlings and Twin Earthlings have no moral dispute. They are merely talking past each other. But this result is highly counterintuitive.

The realist might alternatively claim that moral terms have the same meaning on Earth and Moral Twin Earth, but different referents. This response implies that the same moral judgment may be true for Earthlings and false for Twin Earthlings. Naturalist moral realists cannot tolerate this outcome either, because they claim that moral facts are entailed by natural facts. Moral facts cannot differ where the relevant natural facts are the same.

Let us begin by stipulating that the essential semantic role of the term 'good' in the expression 'good α' is the same on Moral Twin Earth as on Earth, to individuate well-functioning tokens of the type α. The predicate 'good α' refers to the natural property of being a well-functioning α. Now suppose Twin Earthlings claim that good human beings are utility maximizers. Earthlings claim that good human beings are those with the

Aristotelian virtues (as specified, let's say, by one's preferred Aristotelian moral theory, be it Foot's, Hursthouse's, or another one altogether).

We can properly characterize the dispute between the Earthling and the Twin Earthling as genuine, not semantic. Earthling and Twin Earthling uses of moral terms like 'good human being' have the same meaning *and* the same referent as long as humans and twin humans share a characteristic functioning. The account implies that at least one disputant has mistaken beliefs about the referent of the term 'good human being' because the referent is fixed by characteristic human functioning. The referent of the term 'good human being' will be the set of humans who exemplify characteristic human functioning, whatever that turns out to be. If we assume that characteristic human functioning is the same on Earth and Moral Twin Earth (an assumption defended below), the referent of 'good human being' is the same. The view implies that the dispute between the Earthling and Twin Earthling is a genuine dispute about moral beliefs and that at least one disputant has false beliefs. The Twin Earthling is merely mistaken (let us assume) in his belief that the term 'good human being' refers to the set of humans that are unconstrained utility maximizers.

An analogy can clarify this claim. Suppose that the Earthlings and Moral Twin Earthling engage in an ethological dispute. While the natural facts about lions are the same on Earth and Moral Twin Earth, Twin Earthling ethologists claim that good lions have weak jaws and dull claws and Earthling ethologists claim that they have strong jaws and sharp claws.

The referent of the Twin Earthling term 'good lion' is the same as the referent of the Earthling term 'good lion' because lion physiology and behavior is the same (by stipulation) on Earth and Twin Earth. At least one ethologist has false beliefs about the referent of the term 'good lion.'

One implication of the semantics I have sketched is that a speaker can grasp the meaning of the term 'good' in the expression 'good α' but still have mistaken beliefs about the referent of the term 'good α' because she has mistaken beliefs about the characteristic functions of α's or the traits that conduce to their performance. For instance, a speaker can know the meaning of the term 'good' in the expression 'good hygrometer'—e.g., she knows that the term refers to well-functioning hygrometers—yet have mistaken beliefs about the function of hygrometers and thus have mistaken beliefs about the referent of the term 'good' in the expression 'good hygrometer.' If she mistakenly believes that the function of a hygrometer is to cut bread, she will mistakenly believe that the term 'good hygrometer' refers to hygrometers that cut bread well in spite of knowing the meaning of the term 'good.' The referent of the adjective 'good' in the expression 'good α' depends on what kind of thing α is (when α is a functional concept, the referent will depend on what the function of α is). Competent speakers can be mistaken about the natural properties to which the expression 'good α' refers if they have mistaken beliefs about the natural functional properties of α's.

One (or more) of the parties to the ethological dispute may have mistaken beliefs about the characteristic functions of lions or the traits that conduce to their performance, or both. If the disputants know little about lions, they can understand the meaning of the terms 'good' and 'bad' but systematically misapply them with respect to lions.[38] The problem is evidential, not linguistic. Nothing about the dispute between the Earthling and the Twin Earthling implies anti-realism in ethology.

---

[38] This point can be made in another way. Suppose ethologists discover that the mythical chupacabra is a real species. Prior to empirical examination, these ethologists would know what it would *mean* for a chupacabra to be good, even though they have no idea which natural properties are the term's referents.

19

Perhaps the real test case for the view is a Moral Twin Earth on which characteristic twin human functions differ from characteristic human functions. Recall that Horgan and Timmons stipulate that the differences between Earthling and Twin Earthling morals are attributable to a species-wide difference in psychology. Twin humans experience sympathy less readily and guilt more readily than humans. Couldn't this stipulation relativize the reference of moral terms?

The answer depends on the extent of the asymmetry. A slight asymmetry in sentiment is unproblematic. Moderate differences in twin humans' experiences of sympathy and guilt would probably be insufficient to alter their characteristic ends or their characteristic ways of achieving them. After all, not all human beings on Earth experience guilt and sympathy to the same degree.

In any case, Horgan and Timmons cannot push the psychological asymmetry too far. An ethical naturalist claims that moral facts are identical to, or wholly constituted by, natural facts. The natural facts entail the moral facts. The point of Moral Twin Earth is to show how a given naturalist moral theory must accept that the moral facts differ on Earth and Moral Twin Earth in spite of the relevant natural facts being similar. An ethical naturalist will gladly accept that where the natural facts differ, the moral facts can differ as well. If the natural facts about the psychology of twin humans are stipulated to differ drastically from humans', a corresponding asymmetry in moral facts is unproblematic because the stipulation violates the conditions of the thought experiment. If the natural facts about humans and twin humans are not roughly the same, the thought experiment breaks down.

For example, if twin humans had a different evolutionary history that endowed them with "species-wide differences in psychological temperament," then 'good human being' would, in fact, refer to different properties on Moral Twin Earth. Suppose that evolution on

20

Moral Twin Earth landed twin humans with unbounded sympathy; they resemble worker ants endowed with rational agency more than they resemble Earthling humans. Twin humans are constitutionally disposed to impartiality toward kin, indifference toward their own lives and pleasure, and to align their interests entirely with those of their social group. Here the Aristotelian can happily admit that the referent of the term 'good human being' differs on Earth and Moral Twin Earth (whether or not it is appropriate to use the term 'human being' to denote the species on Moral Twin Earth is a question that I will leave aside). Aristotelianism can claim that under those conditions, impersonal benevolence is part of what makes a twin human good.

This kind of relativism is not *objectionable* relativism. It is no more problematic for Aristotelian moral realists than the relativism implied by their view that the terms 'good lion' and 'good elephant' refer to different natural properties on Earth. The Aristotelian claims that *all* evaluations are kind-relative. If natural kind terms like 'lion' or 'elephant' refer to different natural properties, evaluations of such kinds as expressed in terms like 'good lion' and 'good elephant' will refer to different natural properties. Evaluative facts are relative to the kind being evaluated. This sort of relativism is built into the very foundation of Aristotelian ethical naturalism. Indeed, it's the core of the earlier criticism of Moore: nothing is good *simpliciter*; something is only good as a *kind* of thing. When the natural properties of two kinds differ, their respective evaluations differ—e.g., 'good lion' and 'good elephant' refer to different natural properties. But this kind of relativism no more implies anti-realism in ethology than it does in ethics.

[5]

By way of conclusion, then: I have argued that there is reason to think that Aristotelian ethical naturalism meets the two success conditions set by Horgan and Timmons. The view

overcomes extreme indeterminacy of moral reference by disallowing rival natural properties as eligible referents for moral terms. It overcomes objectionable relativism by supplying a semantic story that implies the meaning and referent-properties of moral terms are the same on both Earth and Moral Twin Earth.

My argument does not purport to vindicate neo-Aristotelian ethical naturalism generally; the theory may remain vulnerable to a variety of objections. However, my argument does show that it is not the case that all versions of ethical naturalism fail the test presented by Moral Twin Earth.

References

Boyd, R. (1988). "How to be a Moral Realist," in G. Sayre-McCord (Ed.), *Essays on Moral Realism.* Ithaca: Cornell University Press, pp.181-228.

Boyd, R. (2003). "Finite Beings, Finite Goods: The Semantics, Metaphysics, and Ethics of Naturalist Consequentialism, Part I," *Philosophy and Phenomenological Research*, 66, 505-553.

Brink, D. (1989). *Moral Realism and the Foundations of Ethics.* Cambridge: Cambridge University Press.

Casebeer, W. (2005). *Natural Ethical Facts: Evolution, Connectionism, and Moral Cognition.* Cambridge, MA: The MIT Press; Foot, 2003.

Foot, P. (2003). *Natural Goodness.* Oxford: Clarendon Press.

Geach, P. (1967). "Good and Evil," in P. Foot (Ed.), *Theories of Ethics.* Oxford: Oxford University Press, pp.64-73.

Holland, S. (2001). "Dispositional Theories of Value Meet Moral Twin Earth," *American Philosophical Quarterly*, 38, 177-195.

Horgan, T. and Timmons, M. (1991). "New Wave Moral Realism Meets Moral Twin Earth," *Journal of Philosophical Research*, 16, 447-65.

Horgan, T. and Timmons, M. (1992a). "Troubles on Moral Twin Earth: Moral Queerness Revived," *Synthese*, 92, 221-260.

Horgan, T. and Timmons, M. (1992b). "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived," *Philosophical Papers*, 21, 153-75.

Horgan, T. and Timmons, M. (2000). "Copping Out on Moral Twin Earth." *Synthese*, 124, 139-152.

Horgan, T. and Timmons, M. (2009). "Analytical Moral Functionalism Meets Moral Twin Earth," In I. Ravenscroft (Ed.), *Essays on the Philosophy of Frank Jackson.* Oxford: Blackwell, pp. 221–36.

Hursthouse, R. (1999). *On Virtue Ethics.* Oxford: Clarendon Press.

Jackson, F. (1998). *From Metaphysics to Ethics.* Oxford: Clarendon University Press.

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

Moore, G. (1988). *Principia Ethica.* Amherst, NY: Prometheus Books.

Nussbaum, M. (2000) *Women and Human Development: The Capabilities Approach.* Cambridge: Cambridge University Press.

Putnam, H. (1975). "The Meaning of Meaning," in H. Putnam (Ed.), *Mind, Language, and Reality*. Cambridge: Cambridge University Press, pp.215-271.

Smith, M. (1994). *The Moral Problem*. Oxford: Basil Blackwell.

Sturgeon, N. (1985). "Moral Explanations," in Copp, D. and Zimmerman, D. (Eds.) *Morality, Reason, and Truth*. Totowa, NJ: Rowman and Littlefield, pp. 49-78.

Thomson, J. (1997). "The Right and the Good." *Journal of Philosophy*, 94, 273-298.

Thomson, J. (1994). 'Goodness and Utilitarianism.' *Proceedings and Addresses of the American Philosophical Association*, 67, 7-21.

Thomson, J. (2006). "The Legacy of *Principia*," in Horgan, T. and Timmons, M. (Eds.) *Metaethics after Moore*. Oxford: Oxford University Press, pp. 233-254.

Timmons, M. (1999). *Morality Without Foundations: A Defense of Ethical Contextualism*. New York: Oxford University Press.

Wright, G. (1963). *The Varieties of Goodness*. London: Routledge and K. Paul.