

MORAL TWIN EARTH: THE INTUITIVE ARGUMENT

Heimir Geirsson
Iowa State University

Terence Horgan and Mark Timmons have recently published a series of articles where they attack the new moral realism as developed by Richard Boyd.¹ The new moral realism rests on semantics that uses some of the main ideas of the causal theory of names, especially the idea that kind terms have a “natural” or a synthetic definition, such as ‘water is H₂O.’ By arguing that moral semantics can provide us with synthetic definitions of moral terms it appears that the moral realist has just the tools she needs for naturalistic accommodations of moral realism. For example, the availability of synthetic definitions would help moral realists rebut Moore’s open question argument in a fairly straightforward way.

Against this Horgan and Timmons have argued that our intuitions about the semantics of non-moral language and moral language differ, and that while twin-earth semantic intuitions generate one result in Putnam’s twater case, moral twin-earth fails to generate comparable results for moral terms.² Horgan and Timmons conclude from this that the semantic norms governing the use of natural kind terms differ from the semantic norms governing the use of moral terms. I will argue that Horgan and Timmons’ intuitive moral twin-earth argument fails to derail the new moral realism. Further, I will discuss Boyd’s semantic theory and raise problems for it that do not rely on the use of moral twin-earth.

Moral Semantics

The developments in philosophy of language since the 1970’s have made possible a new wave of moral realism. Boyd’s semantics for moral terms which underlies his moral realism rests on three main strands. First, he maintains that moral terms like ‘good’ have a synthetic definition. With that he implies that just as ‘water’ is not synonymous with some phrase that denotes a natural property, ‘good’ is not synonymous with some phrase that denotes a natural property. Second, he closely follows the lead of the causal theory of reference when claiming that moral terms behave like natural kind terms. Since natural kind terms designate the same substances or properties in all possible worlds in which those substances or properties exist and so are rigid designators, moral terms are rigid designators according to Boyd.

Boyd’s third strand introduces a new wrinkle to causal theories of

reference. While he holds that reference is a matter of there being certain causal connections between the use of moral terms and the relevant natural properties, Boyd's causal theory of names differs from the standard one as developed by Kripke and Putnam. The standard view doesn't say much about how a causal chain is transmitted, but the general view is that once the reference of a name has been fixed then it retains its reference as long as its user intends it to refer to the same object/kind as it did when she acquired the name. For example, once the reference of 'water' has been fixed, 'water' refers to water in all possible worlds, and whenever I use the term 'water' intending the term to have its customary reference it refers to water. But Boyd's account of reference introduces a different account of a causal connection. According to Boyd reference is essentially an epistemic notion and so the relevant causal relations constituting reference are those causal connections involved in knowledge-gathering activities.

Horgan and Timmons formulate Boyd's view as follows:

CRT Causal regulation thesis: For each moral term t (e.g., 'good'), there is a natural property N , such that N and N alone causally regulates the use of t by humans.³

Since CRT allows one to treat moral terms as natural kind terms, they summarize the new moral semantics as follows

CSN Causal semantic naturalism: Each moral term t rigidly designates the natural property N that uniquely causally regulates the use of t by humans.⁴

If CSN is true, then each moral term t should have a synthetically true definition whose definiens characterizes, in purely naturalistic language, the natural property that uniquely regulates the use of t by humans.

Moral Twin Earth; the Intuitive Case

Horgan and Timmons test Boyd's moral semantics by comparing our semantic intuitions about Putnam's water/twater thought experiment and a similarly constructed thought experiment about moral terms. In his thought experiment Putnam introduces twin-earth that is like earth in all respects except that the clear liquid that fills oceans, rivers, and lakes on twin-earth has a different molecular structure than on earth, namely XYZ instead of H₂O on earth. Intuitively, we assume that 'water' in the mouth of earthlings and twin-earthling refers to different substances on earth and twin-earth; namely, 'water' refers to water on earth, and to twater, the watery substance on twin-earth, on twin-earth. So, our intuitive judgment is that the referent, i.e., the semantic meaning, of 'water' is different on earth and twin-earth.⁵

Moral twin-earth is just like earth in most respects. Twin-earthlings behave like earthlings, they make moral judgments and speak Twin-English. If earthlings were to visit twin-earth they would be strongly inclined to think that twin-earthlings use moral terms like earthlings. But there is one crucial difference between earth and twin-earth. When earthlings use moral terms, such as 'good' and 'right', their use of the terms is regulated by certain natural properties distinct from those that regulate the use of twin-earthlings of the same terms. Horgan and Timmons assume that earthlings' use of moral terms is regulated by some consequentialist moral properties, while twin-earthlings' use of moral terms is regulated by some deontological moral properties.

Given the similarities and differences between earth and twin-earth, what is the appropriate way to describe the difference between moral and twin-moral use of moral terms? According to Horgan and Timmons, two options are available. We could say that the differences are analogous to those between earth and twin-earth in Putnam's example, namely that moral terms used by earthlings rigidly designate the natural properties that causally regulate their use while the moral terms used by twin-earthlings rigidly designate the natural properties that causally regulate their use, so the terms refer to different properties on earth and twin-earth. If that is so, then the moral terms used by earthlings and twin-earthlings differ in meaning and are not intertranslatable. The second option is to say that moral and twin-moral terms do not differ in meaning, i.e., that they refer to the same properties, and hence that any moral disagreement that might arise between earthlings and twin-earthlings would be a genuine moral disagreement and not just disagreement in meaning.⁶ Horgan and Timmons claim that the second option is the one that is viable, and that our intuitions therefore do not yield the same results as in Putnam's twater thought experiment, where our intuitions yield the result that the earth and twin-earth terms differ in meaning.

Horgan and Timmons conclude from the moral twin-earth experiment that if CSN were true, and the moral terms in question rigidly designated the natural properties that causally regulate their use, then the moral twin-earth scenario ought to generate intuitions analogous to those generated by Putnam's twin-earth; namely, it should seem natural to claim that there is a difference in meaning and that twin-English moral terms are not translatable by English moral terms. But, they continue, since no such intuitions are generated, the outcome of the thought experiment constitutes strong evidence against CSN.⁷

The Intuition Quandary

There are strong reasons why we should not accept the intuitive moral twin-earth argument. The argument depends on a comparison between our intuitions about moral twin-earth and Putnam's water twin-earth. However, moral twin-earth introduces serious complications that, as I will argue, result in it being unlikely that we can trust our intuitions about the meaning of moral terms. Let's keep in mind that the measuring stick Horgan and Timmons use is Putnam's twin-earth; i.e., they are arguing that our intuitions regarding the reference, i.e. the meaning, of 'water' on Putnam's twin-earth differ from our intuitions regarding the reference, i.e. the meaning, of moral terms on moral twin-earth. We then have the following differences and complications that we do not face with Putnam's twin-earth thought experiment.

1. We are dealing with a single property being changed on Putnam's twin-earth, namely the molecular structure of water, while we are dealing with a change in multiple properties on moral twin-earth, namely changes in the homeostatic properties good, just, right, as well as other moral properties.⁸
2. Further, it is quite likely that we will have a hard time taking a good and hard objective look at our intuitive judgment regarding moral properties as we can when it comes to the molecular structure of water, the reason being that the moral properties are an integral part of us that influence many of our most important decisions and just about every aspect of our lives.
3. As we will see in the following section, a close look reveals that Boyd's account of reference is significantly different from Putnam's account of reference and that the two accounts give us different results across possible worlds. One advocates direct designation, the other causal regulation. Kripke and Putnam are direct designation theorists. It is a causal chain to the baptized object that secures a name's reference, and once we have baptized an individual the name refers to that individual regardless of how she grows, develops, or changes.⁹ In the case of substances, once the reference of the term is fixed, then the term denotes that kind regardless of the phenomenal properties the substance might have. In the case of properties Kripke says, e.g., that heat is something we have identified as giving us a certain sensation which we call 'the sensation of heat,' and that we use this sensation to identify heat, subsequently fixing the reference of 'heat'.¹⁰ Even though we used a certain

sensation when we fixed the reference of ‘heat’ we should not identify heat with whatever causes that sensation, for there are possible worlds in which we are not sensitive to heat. In contrast to Kripke and Putnam, reference for Boyd is an epistemic notion where reference is tracked via causal properties that regulate our use of the terms. We thus have a situation where we are using one semantic theory when working with Putnam’s twin-earth, and a second and significantly different semantic theory when we turn our attention to Boyd. Still, Horgan and Timmons assume in their thought experiment that Boyd’s theory gives us the same results across possible worlds as does Putnam’s theory.

4. Putnam gives us a thought experiment in semantically neutral terms where he elicits our semantic intuitions about the reference of ‘water’. Horgan and Timmons give us a semantically loaded thought experiment where it is, e.g., built into the thought experiment that the use of moral terms is “causally regulated” in one way on earth, and in a different way on moral twin-earth. Given Boyd’s view, “causal regulation” is a semantically loaded term, since the reference of kind terms is determined by causal regulation.
5. Putnam elicits our intuitions about the reference of ‘water’ and concludes that we regard it as referring to different substances on earth and twin-earth. Horgan and Timmons, when presenting what we can conclude from their thought experiment, include not just reference but also the possibility of moral disagreement. It is thus not clear whether they are using their thought experiment to elicit our intuitions about the reference of moral terms, or our intuitions about whether or not earthlings and twin-earthlings can have moral disagreements.¹¹ For example, my intuition regarding the reference of moral terms on earth and twin-earth is that they do refer to different properties, and my intuition regarding the possibility of moral disagreement is that earthlings and twin-earthlings can have moral disagreement. But this is a result that Horgan and Timmons seem to have a hard time dealing with, for the following reason.
6. Horgan and Timmons assume that if moral terms refer to different properties then earthlings and twin-earthlings can not have moral disagreements. Not only does that assumption further complicate their case, it needs to be *argued* for, and once we need argumentative support we are leaving the realm of intuitions.¹² Once again we have a serious departure from Putnam’s carefully constructed thought experiment.

On the assumption that the reliability of our intuitions diminishes in proportion to the complexity of the case at hand, we should conclude that there is too much going on in the moral twin-earth thought experiment for us to be able to reliably consult our intuitions about the reference of moral terms. What we need is not intuition but analysis.

Boyd's Semantics

Since the intuitive moral twin-earth argument fails to derail semantic moral realism the irrealist has to look for another point of attack. The most direct way to attack semantic moral realism is to question the semantic foundations on which it rests. When doing so I will assume that the moral realists are not devising a semantic theory for moral terms specifically, but that they are instead working with a semantic theory for kind terms that subsequently is applied to moral terms. It is therefore sufficient to show that the theory in general has problems; its application to one kind rather than another will not solve any problem the theory might have.

Horgan and Timmons formulate Boyd's basic semantic view as CSN. However, Boyd should not accept CSN as an appropriate formulation of his basic semantic view. CSN assumes that there is a single natural property, *N*, that causally regulates the use of each moral term by humans, while Boyd emphasizes *homeostatic clusters* of properties as regulating the use of the relevant terms.¹³ Boyd's version has an advantage over the Horgan and Timmons version for the following reason. Suppose that there is no *single* natural property that regulates the use of some term by humans and that instead there are several different properties that regulate the use of the term. Since there is not a single property that regulates the use of the term, the term would fail to refer on Horgan and Timmons' version, but not on Boyd's cluster formulation. CSN therefore needs to be replaced with a formulation that preserves Boyd's cluster properties. We can do that as follows:

BSN *Boyd's semantic naturalism*; A term *t* refers to *k* (kind, property, relation) just in case the properties of *k* uniquely regulate the use of *t* via a causal relation.¹⁴

BSN allows for homeostatic properties to regulate the use of a term and so it avoids the counterexample to CSN.

Because Boyd gives up direct designation in favor of causal regulation of our use of kind terms he leaves his semantic theory open to counterexamples of exactly the kind Kripke and Putnam used to argue against the description theory of kind names. I will look at a couple of the examples that they used to show the inadequacy of the description theory and see how Boyd's view

gives us the same counterintuitive results as does the description theory. The examples will demonstrate that just as a cluster of descriptions can pick out the wrong object, a cluster of properties that regulate our use of a term can pick out the wrong object. The examples will show the crucial difference between the theory of Kripke and Putnam and that of Boyd. The theory of the former has kind names track kinds in virtue of their essential properties, while Boyd's theory has names track kinds in virtue of the properties that regulate our use of the term, and those properties can be, and most often are, phenomenal and not essential properties.

Consider, as Kripke does, a counterfactual situation where fool's gold, or better yet, a substance which counterfeited the phenomenal qualities of gold but lacked its atomic structure, was actually found in all the places that now contain gold. Would we say of this counterfactual situation that gold would not have been an element, since the counterfeit substance is not an element? According to the description theory we should say that gold would not have been an element since, according to the theory, 'gold' refers to the counterfeit substance as well as to gold. But Kripke disagrees, and rightly so, and thinks we should not say of the situation that gold would not have been an element.¹⁵ One should not say that the substance would be gold that lacked the atomic number 79. Instead, we should say that the substance is not gold.

What should Boyd say about the same situation, given BSN? He would have to say that the counterfeit substance is gold, since the counterfeit substance is the substance whose properties causally regulate our use of the term gold. Consequently, Boyd would have us conclude that in the counterfeit situation gold is not an element. So, Kripke (and Putnam) would say, correctly, that the substance is counterfeit for gold, while Boyd and the description theorists would have to say, counterintuitively, that it is gold.

Consider a version of another example of Kripke's.¹⁶ Suppose that we find in some until now unexplored part of the world animals that look just like tigers and have all the phenomenal properties that regulate our use of 'tiger', but turn out to be reptiles when examined more closely. Do we then say that some tigers are reptiles? Since the "reptile tigers" satisfy most of the descriptions that constitute the meaning of 'tiger', the description theory would have us say that the reptiles are tigers, while Kripke says they are not, and rightly so. Boyd would have to say that since the properties of both tigers and the new-found reptiles regulate our use of 'tiger', the reptiles are tigers. Again, Boyd's semantic theory gives us the same counterintuitive result as does the description theory.

Boyd can offer two responses to the above. First, he can say that most, but not all of the properties that regulate our use of ‘tiger’ also belong to the reptiles and that ‘tiger’ refers to the kind that has most of the properties that regulate our use of the term. But that will not do, since we can encounter three-legged tigers, albino tigers, or tigers that have lost their tail, in which case they lack many of the properties that regulate our use of ‘tiger’ but still are tigers. Second, he can point out that it is an a posteriori question just which properties belong in the homeostatic definition of kind terms, such as ‘tiger’, and that once we determine which properties belong in the homeostatic definition we can settle the question of the reference of ‘tiger’. But this response fails to recognize that it is not the properties that belong to the homeostatic definition of a kind term that determine reference for Boyd; instead it is the properties that causally regulate our use of the term. If the phenomenal properties regulate our use of a kind term, then the phenomenal properties are the properties that are causally relevant for reference.

The problems that Boyd’s semantic view encounters are very similar to the problems that the description theory of names encountered at the hands of the direct designation theorists. The description theory held that a name refers to the object, or the kind of object, that best satisfied the cluster of descriptions that constituted the meaning of the term. As the direct designation theorists pointed out, this view led to names designating the “wrong” things – e.g., counterfeit gold instead of gold and “reptile tigers” instead of tigers. We are seeing that Boyd’s semantic view suffers from some of the same defects as the description theory of names. Just as a cluster of descriptions can pick out the wrong object, a cluster of properties that regulate our use of a term can pick out the wrong object. The same examples that show that the description theories are semantically inadequate show that Boyd’s causal regulation view is semantically inadequate.

I have argued that CSN should be replaced with the more plausible BSN, which has the additional virtue of having Boyd’s support. While BSN avoids some problems that CSN falls prey to, the examples that show that BSN should be rejected also show that CSN should be rejected.¹⁷ The semantic foundation of Boyd’s semantic moral realism is not strong enough to support moral realism.¹⁸

Notes

¹ “Troubles for New Wave Moral Semantics: The ‘Open Question Argument’ Revived,” *Philosophical Papers*, 1992 (153-175), “Troubles on Moral Twin Earth: Moral Queerness Revived,” *Synthese*, 1992 (221-260), “New Wave Moral Realism Meets Moral Twin Earth,” *Journal of Philosophical Research*, 1990-1991, (447-

472). See also Timmons' *Morality Without Foundations: A Defense of Ethical Contextualism*, (New York: Oxford University Press, 1999) and "On The Epistemic Status of Considered Moral Judgments," *The Southern Journal of Philosophy*, 1990 (supplement). For Boyd's view, see Richard Boyd, "How To Be a Moral Realist," *Essays on Moral Realism*, ed. Geoffrey Sayre-McCord, (Ithaca: Cornell University Press, 1988): 181-228.

² Horgan and Timmons appear to present a substantially different argument as well, where they claim that if moral realism, as presented by Boyd, is true, then earthling and twin earthlings, when meeting, would talk past each other and thus could not have a genuine disagreement about moral issues. I discuss that argument in "Moral Twin Earth and Moral Semantic Realism." What I call here the intuitive argument is, according to Horgan and Timmons, a recipe for arguments against moral realism.

³ Horgan and Timmons, "Troubles for New Wave Moral Semantics," p. 159, Timmons, *Morality Without Foundations: A Defense of Ethical Contextualism*, p. 58.

⁴ Horgan and Timmons, "Troubles for New Wave Moral Semantics," p. 159, Timmons, *Morality Without Foundations*, p. 58.

⁵ We should note that the primary purpose of Putnam's thought experiment is not to elicit our intuitions about the reference of 'water', but rather to motivate the distinction between wide and narrow content.

⁶ See, e.g., Horgan and Timmons, "Troubles for New Wave Moral Semantics," p. 165.

⁷ See, e.g., Timmons, *Morality Without Foundations*, p. 62.

⁸ Suppose we try to eliminate this problem by focusing, e.g., only on 'good'. In that case we nevertheless have a multitude of properties to deal with on moral twin-earth, since the property of being good is a homeostatic property, meaning that it composed of a multitude of properties that come together in the right way.

⁹ For reference of kind names, see H. Putnam, "It Ain't Necessarily So," *Journal of Philosophy*, 1962, pp. 658-671 and "The Meaning of 'Meaning'," *Language, Mind and Knowledge*, ed. K. Gunderson, (Minneapolis: University of Minnesota Press, 1975), S. Kripke, *Naming and Necessity*, (Cambridge: Harvard University Press, 1980), p. 122.

¹⁰ Kripke, p. 131.

¹¹ Or is the thought experiment perhaps supposed to show that even if we settle the reference of moral terms, there is still a substantive question that remains in the moral case, unlike the twater case, and that this show somehow that moral terms, unlike natural kind terms, are evaluative in nature? If so, then the thought experiment is not testing our semantic intuitions about reference, contrary to what Horgan and Timmons claim, and the use of Putnam's thought experiment and the emphasis on our semantic intuitions becomes a bit of a mystery.

¹² I discuss the inference and argue that it is invalid in "Moral Twin-Earth and Moral Semantic Realism," (in progress). The possibility of genuine disagreement

in spite of the parties using terms that don't have the same semantic meaning arises because the parties can nevertheless *use* their terms to refer to the same objects. That is, the terms can have the same speaker meaning although they do not have the same semantic meaning.

¹³ Boyd uses the term 'healthy' as an illustration of homeostatic cluster definition, although he does not discuss whether or not 'health' is a full-blown kind term.

¹⁴ Boyd, p. 195.

¹⁵ Kripke, p. 124.

¹⁶ Kripke, p. 120.

¹⁷ Instead of finding a cluster of properties that regulate our use of a term, we would need to find a single property that does so. We then set up a counterfactual situation where a different substance has that property (as, e.g., the counterfeit gold). The problem arises in the same way as before.

¹⁸ I thank Robert Feleppa, Margaret Holmgren, Mark van Roojen, and Peter Vranas for their comments on drafts of this paper.