

## *Moral Twin Earth, Intuitions, and Kind Terms*

HEIMIR GEIRSSON

*Department of Philosophy and Religious Studies,  
Iowa State University*

*Horgan and Timmons, with their Moral Twin Earth arguments, argue that the new moral realism falls prey to either objectionable relativism or referential indeterminacy. The Moral Twin Earth thought experiment on which the arguments are based relies in crucial ways on the use of intuitions. First, it builds on Putnam's well-known Twin Earth example and the conclusions drawn from that about the meaning of kind names. Further, it relies on the intuition that were Earthers and Twin Earthers to meet, they would be able to have genuine moral disagreements. I will argue that the similarities with Putnam's thought experiment are questionable and so the reliance on Putnam-like intuitions is questionable. I will then further argue that even if we accept the intuitions that Horgan and Timmons rely on, the anti-realist conclusion is not warranted due to there being more to the meaning of kind terms than the argument assumes. Once we develop the meaning of kind terms further we can acknowledge both that Earthers and Twin Earthers refer to different properties with their moral terms, and that in spite of that they can have a substantive disagreement due to a shared meaning component.*

**Keywords:** Moral twin-earth, intuitions, disagreements, kind names, moral realism.

Terence Horgan and Mark Timmons have written a number of articles where they use their Moral Twin Earth thought experiment to attack the new moral realism (Horgan & Timmons 1990).<sup>1</sup> The new moral realism is based on advances made in the philosophy of language. Suppose, the argument goes, that a causal theory of reference of the kind advanced by Saul Kripke and Hilary Putnam, to name two of the usual suspects, is true. Suppose further that moral terms, such as 'good', 'right', 'kind', and 'just' are kind terms which reference is determined causally and

<sup>1</sup> Later articles include (Horgan & Timmons, 1992a, 1992b, 2000, 2009).

that they are rigid designators so that they refer to the same kinds/objects in all worlds in which they exist. Given this, it seems that the new moral realist has at her disposal synthetic definitions of moral terms. Just as we can say that “water is  $H_2O$ ” is necessary and a posteriori, it seems that we can now provide similar definitions for moral terms, namely definitions that identify moral properties with natural properties. Such definitions would be necessarily true and at the same time only knowable a posteriori. This way of approaching moral realism, namely through the advances in philosophy of language and the causal theory of reference, should provide the new moral realist with ammunition to counter some of the traditional objections against moral naturalism.

The Moral Twin Earth thought experiment relies in crucial ways on the use of intuitions. First, as Horgan and Timmons frequently point out, it builds on Putnam’s well-known Twin Earth thought experiment and the conclusions drawn from that thought experiment about the meaning of kind names. Further, it relies on the intuition that were Earthers and Twin Earthers to meet, they would be able to have genuine moral disagreements.

I will argue that the dissimilarities with Putnam’s thought experiment are significant and so the reliance on Putnam-like intuitions is questionable.<sup>2</sup> I will then further argue that even if we accept the intuitions that Horgan and Timmons rely on, that is, even if we address the thought experiment and arguments on their own and apart from comparisons with Putnam’s example, the anti-realist conclusion is not warranted. There are two main reasons for this. First, there is more to the meaning of kind terms than the argument assumes. That is, the basic intuitions that Twin Earth scenarios provide regarding the meaning of kind terms leave us with an incomplete account of meaning. Once we develop the meaning of kind terms further we can acknowledge that Earthers and Twin Earthers refer to different properties with their moral terms, and in spite of that they can have a substantive disagreement due to a shared meaning component. Second, Horgan and Timmons rely on the intuition that were Earthers and Twin Earthers to meet they would have genuine moral disagreement. Horgan and Timmons rely on that intuition when they argue that the meaning of the relevant terms on Earth and Twin Earth must be the same. I will argue

<sup>2</sup> My objection will not rely on the differences that are developed in (Laurence, Margolis, Dawson 1999: 135–165) They argue that our intuitions on Putnam’s Twin Earth and Moral Twin Earth differ in three main respect. First, Putnam’s example deals with a fictional example while Moral Twin Earth deals with existing competing moral accounts. Second, given the similarities between Earthers and Twin Earthers they claim that there is a reason to believe that the moral properties on earth are instantiated on Twin Earth, and vice versa. In contrast, there is no XYZ on earth. Third, they point out that the relevant moral properties are functional, while Putnam’s example does not deal with functional properties. Michael Rubin discusses these arguments and argues that they fail to introduce significant complications for our intuitive judgment regarding Moral Twin Earth. See (Rubin 2008).

that we can accept that they have genuine moral disagreement while at the same time claim that the terms they use have different referents and so different semantic meaning. That is, having genuine disagreements does not require that the semantic meaning, or the reference of the terms used in the debate be the same.

When Horgan and Timmons first presented their thought experiment and the arguments based on it they targeted specifically Richard Boyd and his causal account of reference. They have since stated that their thought experiment is a generic recipe that can be applied to virtually any metaphysically naturalistic version of moral realism (Horgan & Timmons 2009). While my concern in the bulk of the paper will not be with Boyd's account of reference, I do discuss his account and argue that while his account is broadly causal, it differs significantly from the Kripke/Putnam model. Specifically, I will argue that Boyd's causal account does not accommodate rigid designation and thus does not provide the framework needed for new wave moral realism to succeed.<sup>3</sup>

### *Causal reference and rigid designation*

The description theory of reference holds that the reference of a name is mediated via content. According to some of the more influential versions of the theory a given name has a descriptive meaning, and the name refers to whatever object best satisfies the majority of the descriptions that comprise the meaning of the relevant name. For example, the name 'Thales' has a meaning, and whatever object best satisfies the meaning is Thales. If, for example, the meaning of 'Thales' is "the Greek philosopher who held that all is water," then the name refers to the Greek philosopher who held that all is water, whoever that was.

Against the description theory Kripke, Donnellan, Barcan-Marcus, Putnam, and Kaplan, to name a few, advanced the causal theory of reference. One of the main accomplishments of the causal theory of reference was to separate the connection between content and reference, showing that a mode of presentation does not determine reference. Advocates of the causal theory of reference argued that a name is connected to an object via an initial act of baptism, and the object named is the semantic meaning of the name. When the name is passed from one language user to the next it retains its reference as long as the language users intend to use the name with the same reference it had when passed to them. As an example, there is a causal chain that connects our use of 'Thales' to the initial baptism of Thales, thus maintaining the reference of the name. Whether or not the person at the tail end of the chain held that all is water is inconsequential for the reference of the name to succeed. Instead of reference being mediated through content the name refers to the person at the end of the causal chain.

<sup>3</sup> My arguments here regarding Boyd will be brief. For more extended arguments see (Geirsson 2005).

An important feature of the causal theory of names is that names are not only direct, but also rigid designators, meaning that they refer to the same objects in all possible worlds (in which they exist). Rigidity helps explain how it is that true identity statements, such as “Samuel Clemens is Mark Twain,” are necessarily true, if true. If ‘Samuel Clemens’ and ‘Mark Twain’ are rigid designators, then they designate the same object in all possible worlds (in which Clemens exists). Given that, there will be no worlds in which Clemens is not Twain and so it is necessarily true that Clemens is Twain.

The causal theory of reference has also been used to argue that natural kind terms refer directly and are rigid designators. Suppose I decide to call a type of liquid ‘water’ and then find out that this type of liquid has an atomic structure  $H_2O$ . Since ‘water’ is a rigid designator that refers to the same type of material in all possible worlds in which it exists, ‘water’ refers to water in all possible worlds (in which water exists). It might be the case that water has different phenomenal properties in different worlds. It might, for example, be green in some worlds. But just as it was contingently true that Thales held that all is water, it is contingently true that water has the phenomenal properties it has, such as being a clear liquid in this world and a green liquid in some possible world. What makes water water is its atomic structure, not its phenomenal qualities. Since water has the atomic structure  $H_2O$ , it has that atomic structure in all possible worlds. That is, it is a necessary truth that water is  $H_2O$ . Nevertheless, it would be a mistake to say that it is true in virtue of the meaning of the words that water is  $H_2O$ , for it took substantial research to uncover the fact that water is  $H_2O$ .

Notice that a requirement for names being rigid designators is that their reference is *not* mediated via content that needs to be satisfied in a way specified by the description theory of reference. If the reference of a name is mediated via some content, as description theories of reference suggest, then the content determining reference leads to reference being unstable across worlds. For example, if the meaning of ‘water’ is given as a disjunction of descriptions that describe its phenomenal properties on earth, then a liquid that satisfies those descriptions on Twin Earth would be the referent of ‘water’ in spite of the liquid not being  $H_2O$ . That is, the name ‘water’ would not be rigid. Similarly, if the reference of ‘water’ is somehow regulated by its phenomenal or functional properties such that its reference (causally) tracks the relevant phenomenal or functional properties, then, as I will argue, a different substance with identical phenomenal or functional properties would be the referent of the term in a world where such a substance exists. Hence, the name ‘water’ would not be rigid.

Hilary Putnam’s Twin Earth thought experiment provides some of the most influential insights on the meaning of kind names. Horgan and Timmons frequently invoke Putnam’s thought experiment. They want to produce a Moral Twin Earth example that is such that a “reflection on this scenario generates intuitive judgments that are compa-

rable to those concerning Putnam's original scenario" (Horgan & Timmons 2009: 5). Due to the role Putnam's Twin Earth plays for Horgan and Timmons it is instructive to look more closely at the Twin Earth thought experiment. A close look will show that Putnam's thought experiment, as he uses it, is of no help to Horgan and Timmons. Instead a comparison is detrimental to Horgan's and Timmons' project.

### *Putnam's Twin Earth*

Putnam describes Twin Earth as follows:

Twin Earth is very much like Earth; in fact, people on Twin Earth even speak *English*. In fact, apart from the difference we shall specify in our science-fiction examples, the reader may suppose that Twin Earth is *exactly* like Earth. He may even suppose that he has a *Doppelgänger*—an identical copy—on Twin Earth, if he wishes, although my stories will not depend on this.

Although some of the people on Twin Earth...speak English, there are, not surprisingly, a few tiny differences which we will now describe between the dialects of English spoken on Twin Earth and Standard English. These differences themselves depend on some of the peculiarities of Twin Earth.

One of the peculiarities of Twin Earth is that 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 it 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. (Putnam 1975: 223)

There are four points or lessons that I want to draw from Putnam's Twin Earth thought experiment and his use of it. All four points are relevant when discussing Moral Twin Earth.

*Point 1:* The issue of whether the names 'water<sub>e</sub>' and 'water<sub>t</sub>' have the same reference does not arise in Putnam's Twin Earth thought experiment. Instead, it is assumed that the two names differ in reference, namely 'water<sub>e</sub>' refers to  $H_2O$  while 'water<sub>t</sub>' refers to XYZ. Putnam makes this clear throughout his writing, including in the following when he discusses what happens when Earthers who visit Twin Earth discover that the water-like liquid on Twin Earth is XYZ.

If a spaceship from Earth ever visits Twin Earth, then the supposition at first will be that 'water' has the same meaning on Earth and on Twin Earth. This supposition will be corrected when it is discovered that 'water' on Twin Earth is XYZ, and the Earthian spaceship will report somewhat as follows: 'On Twin Earth the word "water" means XYZ.' (Putnam 1975: 223)

Putnam goes on and states that "there is no problem about the extension of the term 'water'." (Putnam, 1975: 224) What we call 'water' simply is not water on Twin Earth, and what they call 'water' simply is not water on Earth.

*Point 2.* The referent of a kind name does not constitute its full meaning according to Putnam. The referent of a kind name constitutes

a *part* of the meaning of a kind name, but there is more to their meaning than what they denote. Putnam makes this clear when he writes “...note that although ‘means’ does mean something like *has as extension* in this [Twin Earth] example, one would *not* say

‘On Twin Earth the meaning of the word “water” is XYZ.’

Unless, possibly, the fact that ‘water is XYZ’ was known to every adult speaker of English on Twin Earth.” (Putnam 1975: 224)

Given that many followers of the causal theory of names have argued that the meaning of names of individual objects is their referent, Putnam’s use of the ‘means’ and ‘meaning’ is unfortunate, as it sometimes suggests that Putnam is claiming that the meaning of a kind name is its referent. But that is not so. Other factors constitute parts of the meaning of kind names on Putnam’s account. In the case of water, the meaning of ‘water’ includes *syntactic markers*, such as the name being a mass noun, *semantic markers*, such as water being a natural kind and a liquid, *stereotypes*, such as water being colorless, tasteless, and finally its *extension*, namely H<sub>2</sub>O (Putnam 1975: 269).

*Point 3.* Psychological content does not determine reference. Putnam asks us to consider Oscar<sub>1</sub>, a typical Earthian who lived at around 1750, and Oscar<sub>2</sub>, a typical Twin Earthian who lived, of course, at the same time. At this time the typical Earthian did not know that water consisted of hydrogen and oxygen, and the typical Twin Earthian did not know that water<sub>t</sub> consisted of XYZ. Suppose further that there is no belief that Oscar<sub>1</sub> has about water that Oscar<sub>2</sub> does not also have. The two of them are exact duplicates in feelings, thoughts, interior monologue, etc. Nevertheless, as Putnam points out, the extension (reference) of ‘water’ at that time was H<sub>2</sub>O on Earth and XYZ on Twin Earth. That is, psychological content, which is identical in Oscar<sub>1</sub> and Oscar<sub>2</sub> when it comes to water, does not determine reference. (Putnam 1975: 224)

*Point 4.* Kind terms are rigid designators. That is, a kind term refers to the same kind in all possible worlds in which the kind exists.<sup>4</sup> Simple as it is, this lesson is very important. It is rigid designation that provides for the possibility of necessary a posteriori truths, or synthetic definitions. Rigidity provides kind terms with the stability needed to make the relevant statements true in all possible worlds, or necessarily true. Without rigidity synthetic definitions are not possible and so without rigidity new-wave moral realism is not possible. Any semantic theory that gives up rigidity takes away any hope for synthetic definitions, which are the backbone of new-wave moral realism.

### *Moral Twin Earth*

Horgan and Timmons claim that the Moral Twin Earth thought experiment crucially depends on a key assumption about semantic intuitions:

<sup>4</sup> See for example (Putnam 1975: 231).

SI: Competent users of language have an intuitive mastery of semantic norms.

They frequently cite Putnam's Twin Earth and the intuitive agreement we have that 'water' on Earth and Twin Earth refer to different substances as an example of the intuitive mastery that we have of semantic norms. They clearly intend their thought experiment to closely resemble Putnam's thought experiment. In particular, they want their thought experiment to test the same intuitive judgments as does Putnam's Twin Earth example.

Moral Twin Earth is very much like Earth. The geography and surroundings are the same, and the Twin Earthers who live in Twin Australia and Twin U.S.A. speak Twin English which sounds the same as it does on earth. It is of particular importance that Moral Twin Earthers use moral vocabulary in much the same way as Earthers. The English speaking Moral Twin Earthers use terms like 'good', 'right', 'bad', and 'wrong' when evaluating acts and institutions. In particular, these terms are used to reason about the wellbeing of the population on Moral Twin Earth. Were a group of explorers from Moral Earth to visit Twin Earth then they would be inclined to accept the natives moral terms as identical to their own. So similar is their use.

In spite of all the similarities between Earth and Moral Twin Earth there is a crucial difference between the two. Upon investigation into Twin Earthers moral discourse and practice we find that Moral Twin Earthers use of moral terms is causally regulated by some deontological moral properties and so Moral Twin Earthers converge to deontological morality. Earthers use of moral terms, on the other hand, is causally regulated by consequentialist moral properties and so Earthers converge to consequentialist morality. Moral Twin Earthers moral theory is best systematized with a deontological theory  $T^d$  while Earthers moral theory is best systematized with a consequentialistic theory  $T^c$ . In spite of the different theories Horgan and Timmons suggest that the two theories are similar enough so that moral discourse in the two places operates in much the same manner. The differences in the moralities of Earthers and Moral Twin Earthers, Horgan and Timmons suggest, are in part due to certain species-wide psychological temperament that differs between the two. For example, they suggest, Moral Twin Earthers might experience certain sentiments, such as guilt, more frequently and more intensely than Earthers, and they might experience sympathy to a lesser degree than Earthers (Horgan & Timmons 2009: 7). However, the main reason for the difference in morality is that on Earth peoples use of moral terms is causally regulated by consequentialistic moral properties while on Moral Twin Earth peoples use of moral terms is causally regulated by deontological moral properties.

Given this description of Moral Twin Earth, what is the appropriate way to describe the differences between moral and twin-moral uses of

moral terms, such as ‘good’, ‘right’, ‘bad’, ‘fair’, etc.? Horgan and Timmons claim that two options are available. On the one hand, one could say that the differences we encounter with Earth and Moral Twin Earth are analogous to those between Putnam’s Earth and Twin Earth, namely that moral terms used by Earthers and moral terms used by Moral Twin Earthers refer to different moral properties. If that is so, then moral terms on Earth and Moral Twin Earth differ in meaning and are not intertranslatable. On the other hand, one could say that moral and twin-moral terms do not differ in meaning. If that is so, then any apparent moral disagreement between Earthers and Moral Twin Earthers would be a *genuine* disagreement. That is, it would be a disagreement in moral belief and in moral theory rather than a mere difference in meaning. Horgan and Timmons submit that the natural and plausible alternative is the second one. That is, Horgan and Timmons claim that the intuitions generated by Moral Twin Earth differ from the intuitions generated by Putnam’s Twin Earth thought experiment as in the former, and not the latter, we want to say that the relevant terms have the same meaning.<sup>5</sup>

### *The main differences between Putnam’s Twin Earth and Moral Twin Earth*

When Horgan and Timmons presented Moral Twin Earth they made references to Putnam’s original Twin Earth example and claimed that the intuitions that we draw upon with the Moral Twin Earth thought experiment are, in essence, the same as those elicited by Putnam’s thought experiment. My contention is that they are wrong on this. Let us look at the various lessons drawn from Putnam’s case and find out whether they apply to the Moral Twin Earth scenario.

*Point 1 applied.* In Putnam’s example the issue of what the relevant terms refer to did not arise. It was assumed all along that ‘water<sub>e</sub>’ and ‘water<sub>t</sub>’ refer to different substances. In the Moral Twin Earth example the issue of disagreement and what it means for reference is a focal point. Horgan and Timmons initially assume that moral terms are causally regulated by different moral properties on Earth and Moral Twin Earth. However, given our intuitive judgment that Earthers and Moral Twin Earthers can have a genuine moral disagreement we conclude that moral terms mean the same on Earth and Moral Twin Earth, and hence that the initial assumption about causal regulation (reference) is false. Since it is not the case that moral terms are causally regulated by the different moral properties on Earth and Moral Twin Earth, the relevant moral properties are not a part of the meaning of moral terms. The argument, which we can call The Argument from Disagreement, can be stated as follows:

<sup>5</sup> See for example (Horgan & Timmons 2009: 9).

1. If 'water<sub>e</sub>' and water<sub>t</sub>' refer to different kinds/properties, then Earthers and Moral Twin Earthers would not be able to have genuine moral disagreements.
2. Earthers and Moral Twin Earthers can have genuine moral disagreements.
3. So, 'water<sub>e</sub>' and water<sub>t</sub>' do not refer to different kinds/properties.

It is worth noting that the intuitions appealed to here by Horgan and Timmons have nothing to do with our competent grasp of semantic norms. That is, they do not appeal to intuitions about reference. Instead, after consulting our intuitions regarding *disagreements* Horgan and Timmons infer that 'water<sub>e</sub>' and water<sub>t</sub>' do not refer to different kinds or properties. Putnam, on the other hand, makes no use of intuitions regarding disagreements in his Twin Earth thought experiment and he does not conclude anything about reference from the issue of disagreement. There is therefore a clear disanalogy here between Putnam's use of Twin Earth and Horgan and Timmons' use of Moral Twin Earth. Even more seriously, as I will argue in the next section, The Argument from Disagreements is not sound. Aligning the issues of reference and disagreements, as we will then see, is a mistake that we find in Horgan and Timmons' Moral Twin Earth thought experiment but not in Putnam's Twin Earth thought experiment.

*Point 2 applied.* As Putnam made very clear in his discussion of the Twin Earth thought experiment, the referent of a kind term does not constitute the term's full meaning. Instead the referent is a part of its meaning. Other elements of the meaning of 'water' include it being a mass noun, it being a natural kind and a liquid, and it being colorless and tasteless. Notice that 'water<sub>e</sub>' and 'water<sub>t</sub>' share much, and perhaps most of their meaning. The reference of the two differ, but the remaining elements of their meaning are the same. Like H<sub>2</sub>O, XYZ is a natural kind, it is colorless and tasteless, it is a liquid, and 'water<sub>t</sub>' is a mass noun as is 'water<sub>e</sub>'. Given how much of the meaning of 'water<sub>e</sub>' and 'water<sub>t</sub>' is shared, we can probably say that *to a large extent* they have the same meaning although their reference differs.

Once we recognize that there is a large shared meaning between the two terms, then that opens the door for genuine disagreements to arise about water<sub>t</sub> between Earthers who visit Twin Earth and Twin Earthers even if the two names have different reference. The kind terms need not have exactly the same meaning in order for a genuine disagreement to arise. It suffices that there is a significant overlap of meaning.

In order to see that a significant disagreement can arise in spite of the terms involved not sharing their full meaning consider moral disagreements that we have on Earth. A virtue ethicist, a deontologist, a moral pluralist, and someone who accepts a teleological approach understand moral terms in somewhat different ways. Each theory implies an understanding of key moral concepts and the differences in understanding between, for example, a deontologist and a teleologist

are fairly significant. For example, one places the concept *good* and the other the concept *right* at the foundation of her moral reasoning, and one evaluates acts by their consequences while the other focuses more on duties. Consequently, it is reasonable to claim that when a consequentialist says that an act is right she means something different by the term ‘right’ than does a deontologist who makes the same utterance. In spite of these differences in moral theories on Earth and the corresponding differences in the meaning of moral terms we still manage to have significant and genuine moral disagreement.

What is more pertinent to the Moral Twin Earth thought experiment is the possibility of a disagreement arising in spite of the names used in the debate referring to different substances or properties in different worlds, that is, without the names having the same semantic meaning. The way in which such disagreements can arise relies on speakers being able to *use* words to refer to something other than they denote.<sup>6</sup> Keith Donnellan and Saul Kripke provided numerous examples of such divergence between semantic reference and speaker reference.<sup>7</sup> Someone can, for example, use the description “the man drinking a martini” to refer to someone she, mistakenly, believes is drinking a martini. In fact the man has water in his glass. (Donnellan, 1966) Same goes for names. If I mistake Peter for Paul, then I can use the name ‘Paul’ to refer to Peter. My respondents might understand who I am referring to and so we might have a discussion about the person, namely Peter, while all the time using the name ‘Paul’. Similarly, an Earther visiting Twin Earth can easily use ‘water’ to refer to water<sub>t</sub> and so an Earther and a Twin Earther can use ‘water’ to refer to the same substance. An Earther visiting Twin Earth might look at a lake saying “the water in the lake is pristine,” thus referring to the water<sub>t</sub> in the lake. In the ensuing conversation with a Twin Earther both may succeed in referring to the liquid in the lake, namely water<sub>t</sub>. The Twin Earther might even disagree with the Earther about the water<sub>t</sub>’s clarity and so a genuine disagreement might arise. It helps here how easy it is to mistake water<sub>t</sub> for water<sub>e</sub>. Water<sub>e</sub> and water<sub>t</sub> share a number of properties, including all of their phenomenal properties, and only a few of their known properties, given Putnam’s story, depend on the chemical composition of the two kinds.

Given the identical phenomenal properties of the two kinds and the identical and important roles that the two kinds play on their respective earths, it is easy to see Earthers and Twin Earthers have a heated discussion about, for example, how best to limit access to groundwater, how to clean polluted streams, how to control runoff, how to use tides to produce electricity, etc. They might also disagree about the aesthetic qualities of the moon reflecting in a lake, or how thick the ice has to be before it is safe to walk across it.

<sup>6</sup> I argue for this in greater detail in (Geirsson 2005).

<sup>7</sup> See Donnellan’s and Kripke’s well-known examples (Donnellan 1966, Kripke 1979).

Suppose that after a long argument the Earthers visiting Twin Earth and the Twin Earthers come to an agreement on a brilliant solution about how best to control access to groundwater and how best to deliver it to faraway places. Suppose also that once they have reached this agreement, presented the solution, and received great praise for its originality they find out that water on Earth and Twin Earth are different kinds. Are we now to say that their disagreements and the subsequent solution is void and meaningless because they were all along talking about different substances? Of course not. The disagreements were genuine and the subsequent solution remains brilliant. It does not matter here that water<sub>e</sub> and water<sub>t</sub> are different kinds. The disagreement and the successful subsequent solution reflects the fact that the parties involved were successful in referring to the same substance and so had meaningful disagreements and conversations about that substance. What made the genuine disagreement possible is the fact that the two parties were able to *use* 'water' to refer to water<sub>t</sub> and so they managed to talk about the same substance, namely water<sub>t</sub>.

The point here is simple. Even if two terms refer to different objects/kinds/properties, users of the terms can have genuine disagreements. It is sufficient that the parties *use* their terms to refer to the same objects/kind/property. So, the Argument from Disagreement is not sound.

*Point 3 applied.* The third lesson learned from Putnam's Twin Earth, and arguably the most important point that Putnam makes with his Twin Earth example, is that psychological content does not determine reference. Oscar<sub>1</sub> who lives on Earth and Oscar<sub>2</sub> who is an inhabitant of Twin Earth have the same psychological content and nevertheless their typical utterances of the word 'water' refer to different substances. As with Point 1, the parallel between the two thought experiments breaks down. While Putnam stresses the point that Oscar<sub>1</sub> and Oscar<sub>2</sub> have the same psychological content when it comes to the concept of water while the reference of 'water' nevertheless differs, Horgan and Timmons do no such thing. Instead, Horgan and Timmons specify that moral equilibrium plays a role in the formulation of the relevant moral theories, and hence the formulation of moral concepts. Given that Earthers accept or are guided by a teleological theory while the Twin Earthers accept or are guided by a deontic theory the moral concepts differ regardless of the reference of the relevant moral terms. That is, the psychological content of Earthers and Moral Twin Earthers when it comes to moral concepts is not the same. This is a significant departure from Putnam's Twin Earth.

The departure regarding psychological content from Putnam's thought experiment is significant for two reasons. First, since Horgan and Timmons repeatedly claim to be drawing on the same intuitions in their thought experiment as Putnam does in his, the relevant details in the two thought experiments need to be the same. A departure as serious as this one jeopardizes any meaningful comparison of the two

thought experiments. Second, and more significantly, leaving aside any comparison with Putnam's Twin Earth, the move threatens to undermine the main conclusion that Horgan and Timmons want to draw.

Horgan and Timmons claim that, unlike Putnam's Twin Earth, Moral Twin Earth has us conclude that the relevant kind terms do *not* differ in meaning. Given how they set up the thought experiment it is hard to see how we can conclude that. Because moral equilibrium plays a role in forming the relevant moral theories and concepts, and because it is assumed that Earthers accept a teleological theory while Moral Twin Earthers accept a deontological theory, it is *assumed* from the start that moral terms on Earth and Moral Twin Earth differ in narrow meaning, that is, the part of meaning that resides within one's head. A deontologist has different foundational concepts than does the teleologist, and a deontologist understands the concept of duty, to name one example, in a different way than does the teleologist. Given this it seems clear that Horgan and Timmons cannot conclude, as they do, that moral terms on Earth and Moral Twin Earth have the same meaning.

Horgan and Timmons may respond by pointing out that when they set up their thought experiment they are using 'meaning' as extension or reference. That is, they can claim that the crux of the example has to do with Earthers and Moral Twin Earthers using terms that have to refer to the same kind or properties if a genuine disagreement is to arise. But a response along these lines will not help their cause for the following reason. Consider these alternatives.

A: It is a necessary condition for a genuine disagreement to arise that the referring terms used in the debate refer to the same things.

The requirement set forth in (A) is too strong, as already shown. Instead of (A), the weaker (B) is preferable.

B: It is a necessary condition for a genuine disagreement to arise that the referring terms used in the debate are *used to refer to the same things*.

As we have already seen, terms can be used to refer to things and kinds that the terms themselves do not refer to. Hence the notion of speaker reference (setting it apart from semantic reference). But given (B) it is clear that the Moral Twin Earth thought experiment does not show us that a genuine disagreement can arise only if moral terms refer to the same properties on Earth and Moral Twin Earth, assuming that there are such moral properties.

*Point 4 applied.* Putnam held that kind terms are rigid designators, namely that a given kind term refers to the same kind in all worlds in which that kind exists. Horgan and Timmons argue that moral terms on Earth and Moral Twin Earth have the same meaning. Since they have the same meaning and since, by hypothesis, moral properties on Earth and Moral Twin Earth differ (as our use of moral terms is caus-

ally regulated by different properties on Earth and Moral Twin Earth), moral terms are not rigid designators. Since moral terms are not rigid designators, and since the synthetic definitions of the new moral realism require that moral terms are rigid designators, the new moral realism cannot get off the ground.<sup>8</sup>

If all we can do is point to intuitions regarding the rigidity of kind names, then we would probably have to stop at this point and simply acknowledge that there is a fundamental difference in intuitive insight. But we do not need to stop there. Both parties provide more than intuitive support for their views.

After presenting examples that support an initial view that kind names are rigid designators, Kripke and Putnam proceed to provide additional support. The support they provide consists mostly of arguments showing that the assumption that kind names are rigid designators helps us gain new, deep, and interesting insights into the nature of necessity and the relationships between necessary truths and a priori knowledge. The support consists, for example, of explaining why identity statements are necessarily true if true, and how certain definitions can be at the same time necessarily true and knowable only a posteriori. Additionally, the arguments helped us untangle the many notions associated with the concept of analyticity. One important step here was to make a rather sharp distinction between the metaphysical notions of necessity/contingently and the epistemic notions of a priori/a posteriori. Without these interesting and plausible results, the intuitive insight alone about the rigidity of kind names might not have amounted to much.

Horgan and Timmons do not rely on an intuition about the rigidity and/or reference of moral terms when presenting the main results of their Moral Twin Earth thought experiment. Instead, the intuition they rely on has to do with whether or not Earthers and Moral Twin Earthers can have a genuine disagreement if they meet and start exchanging opinions about moral matters. Once we have the intuition that the two can have a genuine disagreement, Horgan and Timmons *conclude* that the moral terms have the same meaning on Earth as they do on Moral Twin Earth. As has already been shown, it does not follow from the premise that two persons have a genuine disagreement about an object or a property or a kind that the terms themselves used in the debate have to have that object of property or kind as their extension or reference. Other explanations are readily available, including pointing out that the disagreeing parties can *use* their terms to refer to object/property/kind *p* even though the term itself does not refer to *p*, and pointing

<sup>8</sup> We can even reach a stronger conclusion here. If the meaning of moral terms is the same on Earth and Moral Twin Earth, and our use of moral terms is causally regulated by different properties on Earth and Moral Twin Earth, then those properties are not a part of the meaning of moral terms. That is, we get a conclusion that strongly opposes Putnam's point that the referent of a kind term is a part of the meaning of that term.

out that even though the extension of the two terms might differ, they can nevertheless share a substantial amount of their meaning. Given the interesting and plausible results from viewing kind names, in general, as rigid designators, Horgan and Timmons need to do more than they have done so far to establish their conclusion that moral terms are not rigid and so differ in that respect from other kind terms.<sup>9</sup>

### *The Appeal of Moral Twin Earth.*

The main conclusion that Horgan and Timmons draw from the Moral Twin Earth thought experiment is that the meaning of moral terms on Earth and Moral Twin Earth is the same. As a careful comparison between Putnam's Twin Earth and Moral Twin Earth shows, Horgan and Timmons cannot get their conclusion by appealing to the same intuitions as did Putnam in his Twin Earth thought experiment. The differences between the two thought experiments are too great.

Further, we have uncovered that the main inference that Horgan and Timmons rely on, namely the move they make from genuine disagreement to same meaning, relies on a false premise and hence an unsound argument and so their reasoning does not stand on its own. As already pointed out, Horgan and Timmons' conclusion that the meaning of moral terms is the same on Earth and Moral Twin Earth is not directly based on intuition. Instead, the conclusion is inferred from the intuition that Earthers and Moral Twin Earthers would have genuine disagreements about moral issues. We can grant the intuitive insight about them having genuine disagreement. The conclusion that moral terms have the same meaning on Earth and Moral Twin Earth does not follow from that. Hence, we have uncovered that quite apart from the difference between the two thought experiments, the main move that Horgan and Timmons make rests on an unsound argument. Given this the question remains, why do so many accept the conclusion that Horgan and Timmons want to draw from the Moral Twin Earth thought experiment?

One possible explanation is that many, if not most of those involved in the debate assume that the meaning of kind terms is their reference. It is fairly common for causal reference theorists to hold that the meaning of names of individuals is their referent. For example, the meaning of 'Plato' is Plato. But that view is rarely extended to kind names. The *qua* problem, namely the question of what the reference of a name is being fixed to, has forced the inclusion of elements other than the kind itself in the meaning of kind terms, be those descriptive or other

<sup>9</sup> Interestingly some ethicists have suggested that moral terms are either functional terms, or that their reference is guided by some epistemic criteria. For the former, see for example (Brink 2001) For the latter, see for example (van Roojen 2006). Both moves are non-starters for the new moral realist, since they sacrifice the idea of moral terms being rigid designators, and hence give up the possibility of synthetic definitions.

cognitive elements.<sup>10</sup> It is true that if ‘water’ on Earth and Moral Twin Earth have different extensions, then the two terms differ in meaning. But given that the meaning of kind terms is not exhausted by their reference, kind terms might share enough of their meaning so that a genuine disagreement can arise even if, unbeknownst to the disagreeing parties, the terms do not refer to the same kind.

A second possible explanation is that people mistakenly infer that if the terms themselves do not refer to the same kind or properties, then the disagreeing parties are talking about different things, or talking past each other, and so the disagreement is not genuine. Again, this ignores the familiar distinction between speaker reference and semantic reference. That distinction makes it clear that the terms can be *used* by the speakers to refer to objects other than they designate. Since the Earthers and Moral Twin Earthers can use their terms to refer to the same kind they can have a genuine disagreement about it.

A third possible explanation might be that those who are persuaded by Moral Twin Earth are subjectivists about moral values to begin with and are predisposed to find any realist account of ethics implausible. Such predisposition might make them too agreeable to Horgan’s and Timmons’ Moral Twin Earth thought experiment. But as often is the case with thought experiments, the description of Moral Twin Earth is brief and a number of detail are left out. Once we start to think more carefully about the thought experiment we realize that complex issues of reference and meaning cloud the issue and we recognize that the frequently cited Twin Earth thought experiment differs in significant way from the Moral Twin Earth thought experiment and so the intuitions that drive one cannot be applied to the other.

### *The Speculative Nature of New Wave Moral Realism.*

The new wave moral realism started with Richard Boyd’s speculative account of moral properties being homeostatic cluster properties that causally regulate our use of moral terms. (Boyd 1988) The account is speculative because Boyd does not set out to argue that moral realism is true, and he does not argue that moral properties are in fact homeostatic cluster properties or natural properties. Rather, he suggests that the roadmap he provides is one way a realist argument might go. He relies on a causal account of reference and the possibility of synthetic definitions. However, his account of reference, which he now calls *the accommodationist conception of reference* to distinguish it from causal accounts proper, differs significantly from those of the typical causal theorists, such as those of Kripke and Putnam. (Boyd 2010) In particular, he suggests that our use of moral terms is *causally regulated* by moral properties. “[T]he accommodationist conception differs from other causal conceptions of reference...by emphasizing the causal role of

<sup>10</sup> For a useful discussion of these issues see (Stanford & Kitcher 1997).

actual conceptual, descriptive, and intentional practices.” (Boyd 2010: 224) All of these have it in common, Boyd claims, that they play a role in how we understand kind terms, and the role they play is causal since they are relevant to the cognitive and social practices that help determine the meaning of kind terms.

The accommodationist conception of reference with its introduction of causal regulation is a significant departure from the causal account of reference as developed by, for example, Kripke and Putnam. The latter emphasizes our use of names being linked to an object or a kind via a causal chain after the reference has been fixed. Causal regulation, on the other hand, emphasizing actual conceptual, descriptive, and intentional practices, is more likely to track phenomenal properties than kinds. If a term tracks kind *K* in the actual world, and kind *Q* has the phenomenal properties of kind *K* in world *W* and kind *K*, if it exists in *W* has different phenomenal properties than it does in the actual world, then the accommodationist account of reference has us refer to *Q* at *W* when using the term.

How does, for example, Putnam’s use of ‘water’ compare and contrast with Boyd’s accommodationist conception of reference? Consider an Earther who travels to Putnam’s Twin Earth and, unaware of the chemical composition of water<sub>*e*</sub>, uses ‘water’ when communicating with Twin Earthers about the clear liquid in lakes and rivers. His use of ‘water’ while on Twin Earth is causally regulated by water<sub>*e*</sub>. Unable to detect any difference between water<sub>*e*</sub> and water<sub>*t*</sub>, his use of ‘water’ seamlessly fits in with his use of the term prior to arriving at Moral Twin Earth. This becomes even more obvious if we consider this taking place at around 1750, when it is clear that the phenomenal properties of the respective liquids regulate our use of the terms. The visitors conceptual, descriptive, and intentional practices when using the term have not changed, since she believes that the liquid on Moral Twin Earth is the same as it is on Earth. Following Boyd we should then conclude that ‘water’ in the mouth of the Earther now refers to water<sub>*t*</sub>. But this is exactly what Putnam denied in his Twin Earth example. On Putnam’s account the Earther and the Twin Earther refer to different substances. The key difference between the two accounts is that Putnam, as do most causal theorist, claims that kind names are not just causal designators but also *rigid* designators. That is, they refer to the same kind in all possible worlds in which that kind exists. The accommodationist conception of reference, with its causal regulation of use of kind terms, gives up rigidity.<sup>11</sup> The reason is simple. If our use of a kind name is causally regulated in the way Boyd specifies, namely by the relevant cognitive and social practices and with our engagement with the world, then our use of the name tracks properties that govern our use of the name. Those properties are typically the phenomenal and functional properties of the kind. For example, our everyday use of ‘water’ would

<sup>11</sup> I discuss this in greater detail in (Geirsson 2005).

be causally regulated by the phenomenal properties of a liquid, namely look, feel, taste, etc., as well as its various functional roles. Since ‘water’ tracks properties that govern our use, on the accommodationist conception of reference, and since the properties that govern our use belong to different kinds on Earth and Twin Earth, ‘water’ refers to different kinds on Earth and Twin Earth. Rigid designation is not preserved.

Similarly, our use of ‘good’ and other moral terms would be causally regulated by facts about human well-being. The properties that contribute to human well-being might be teleological on Earth and deontological on Moral Twin Earth, in which case moral terms would be causally regulated by different properties on Earth and Moral Twin Earth and so refer to different properties on Earth and Twin Earth. It is this kind of use that Putnam warns against when he claims that we should not mistake such “operational definitions,” namely definitions that depend on our use of the term, for the real meaning of the term (Putnam 1975).

But, the objection might go, any account of reference can be rigidified. Perhaps that is so, but doing so sometimes has consequences that are not wanted. Suppose that we somehow rigidify Boyd’s account of reference. For example, we might index reference to causal regulation at a world so that, for example, the name ‘water’ used by an Earther on Earth refers to what causally regulates her use of the word at the actual world. If we do this, then ‘water’ in the mouth of an Earther who visits Twin Earth refers to  $\text{water}_c$ . However, once we do this then we change the role of causal regulation that Boyd uses. Instead of ‘water’ referring to whatever it is that causally regulates our use of the term, it refers to whatever it is that causally regulates our use of the term *at the actual world*. That is, it is now part of the meaning of ‘water’ that it refers to what causally regulates it in the actual world. And now we can raise a problem for the rigidified account. Suppose there is a Perfect Earth, namely a possible world that, unlike Twin Earth and Moral Twin Earth, is exactly like Earth in every way. My duplicate on Perfect Earth uses ‘water’ and the term refers to what causally regulates the term on Perfect Earth. Consequently, it is a part of the meaning of ‘water’ as used by my twin that it refers to what causally regulates his use of the term at Perfect Earth. But now, contrary to our intuitive judgment, when I think that water is quenching and when by twin on Perfect Earth thinks that water is quenching, then our thoughts have different contents. My thought is partly about the actual world while my twin’s thought is partly about Perfect Earth. That is, even though we are thinking about the same kind or substance our thoughts are different because they are entertained at different worlds.<sup>12</sup> This certainly goes counter to the intuitions that, for example, Putnam relied on when it came to kind terms and so we cannot rigidify Boyd’s account of reference in this way.

<sup>12</sup> The example here is an abbreviated version of Scott Soames’ main argument against rigidified descriptivism in (Soames 2002).

The relevant point of disagreement between the causal account of reference and the accommodationist conception of reference boils down to this. The causal account of reference, as developed by Kripke and Putnam, tracks kinds. Since the causal account of reference tracks kinds, it can support rigid designation. The causal regulation of the accommodationist conception tracks properties that govern our use of the term and so is more likely to track phenomenal or functional properties and our engagement with the world as governed by our social and cognitive practices, meaning that the same term might designate different kinds at different possible worlds. As a result, the accommodationist conception does not support rigid designation.

David Brink's account of names has similar results. He provides a functional account of reference in his attempt to explain the reference of moral terms. For example, the word 'milk', he argues, should be regarded as a functional term. Even though the chemical composition of the white liquid that mammals produce to feed their young differs on Earth and Twin Earth, both are milk, since they serve the same function on the respective earths. Consequently, the word 'milk' refers to the respective liquid in both places even though their chemical composition differs and 'milk' has the same meaning on Earth and Twin Earth. This example paves the way to treat other kind terms, such as moral terms, as being functional terms. However, this account also has the consequence for the new moral realist that reference is not stable across worlds, which is what rigidity requires. 'Water' on earth refers to H<sub>2</sub>O while it refers to XYZ on Twin Earth.<sup>13</sup> However, since the liquid on Twin Earth is functionally equivalent to water on Earth, 'water', on Brink's account, would have the same meaning on Earth and Twin Earth. Not only does this go against Putnam's original insight. This has the results that we lose rigidity, since 'water' now refers to different substances on Earth and Twin Earth. Without rigidity we do not have and synthetic definitions. And without synthetic definitions the new moral realism does not get off the ground.

If we give up causal regulation and functional accounts of names in favor of an account that preserves rigid designation, then the would-be moral realist faces a different kind of a problem, and that is how to fix the reference of a name of an object or a property that might not exist. The causal account of reference faces difficult problems when it comes to non-referring terms and, for all we know today, moral terms may fail to refer. We can talk as if moral properties exist and use moral terms accordingly, just as we can talk as if Santa Claus exists and use the name accordingly. Nevertheless, just as it is questionable whether causal accounts of reference can provide an adequate account of non-referring names, such as 'Santa Claus', it is not clear whether they can provide an account of moral terms. The causal account of reference car-

<sup>13</sup> One might object here and claim that Earth and Moral Twin Earth might not belong to different possible worlds. However, the point still stands. If reference is not stable *in* a world, it surely is not stable *across* worlds.

ries with it a presumption of existence. Such presumption is, as of now, not warranted when it comes to moral terms.

Suppose then that moral properties do exist and that they are homeostatic properties of the kind Boyd suggests. Even then the causal theorist faces difficulties. The causal account of names works well when we are dealing with individual objects or even a kind. But homeostatic cluster properties are neither of those. Instead they are numerous properties that typically go together, or at least most of them. So far little or nothing has been said about how we can individuate and re-identify such property clusters. In the case of water, our predecessors had a criteria for individuation and re-identification. They could name a sample of water and claim that everything that bears the relevant sameness relation to the sample named is water. We have no idea as of how to do the same for moral properties. Until we have a workable criteria, even a rough one, as of how to identify the relevant moral property clusters, it seems that we cannot apply the causal theory of reference to them. If we can, it is incumbent upon the new moral realist to show that we can do that.

In the end, then, the subjectivist does not need to rely on Moral Twin Earth to counter the new moral realist. The subjectivist is better off if she sticks closer to semantics and points out that the prominent Boyd-style new moral realist fails to preserve rigid designation in her account. When rigid designation is gone, so are synthetic definitions, and the latter are crucial to a successful new moral realist account.

## References

- Boyd, R. 1988. "How to Be a Moral Realist." In G. Sayre-McCord (Ed.), *Essays on Moral Realism*. Ithaca: Cornell University Press:181–228.
- \_\_\_\_\_. 2010. "Realism, Natural Kinds, and Philosophical Methods." In H. Beebe & N. Sabbarton-Leary (Eds.). *The Semantics and Metaphysics of Natural Kinds*. New York: Routledge:212–234.
- Brink, D. O. 2001. "Realism, Naturalism, and Moral Semantics." *Social Philosophy and Policy* 18 (2):154–176.
- Donnellan, K. 1966. "Reference and Definite Descriptions." *The Philosophical Review* 75:281–304.
- Geirsson, H. 2005. "Moral Twin-Earth and Semantic Moral Realism." *Erkenntnis* 62:253–278.
- Horgan, T., & Timmons, M. 1990. "New Wave Moral Realism Meets Moral Twin Earth." *Journal of Philosophical Research* XVI:447–465.
- \_\_\_\_\_. 1992a. "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived." *Philosophical Papers* 21:151–173.
- \_\_\_\_\_. 1992b. "Troubles on Moral Twin Earth: Moral Queerness Revived." *Synthese* 66 (92):221–260.
- \_\_\_\_\_. 2000. "Copping Out on Moral Twin Earth." *Synthese* 124:139–152.
- \_\_\_\_\_. 2009. "Analytical Moral Functionalism Meets Moral Twin Earth." In I. Ravenscroft (Ed.). *Minds, Ethics, and Conditionals: Themes from the Philosophy of Frank Jackson*. Oxford: Oxford University Press.

- Kripke, S. 1979. "Speaker's Reference and Semantic Reference." In P. A. French, T. E. Uehling & H. K. Wettstein (Eds.). *Contemporary Perspectives in the Philosophy of Language*. Minneapolis: University of Minnesota Press: 6–27.
- Laurence, S., Margolis, E., Dawson, A. 1999. "Moral Realism and Twin Earth." *Facta Philosophica*, 1:135–165.
- Putnam, H. 1975. "The meaning of 'meaning'." In *Mind, Language and Reality: Philosophical Papers Vol. 2*. Cambridge: Cambridge University Press:215–271.
- van Roojen, M. 2006. "Knowing Enough To Disagree: A New Response to the Moral Twin Earth Argument" *Oxford Studies In Metaethics, Vol. 1*. Oxford: Oxford University Press: 161–193
- Rubin, M. 2008. "Sound Intuitions on Moral Twin Earth." *Philosophical Studies* 139:307–327.
- Soames, S. 2002. *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. New York: Oxford University Press.
- Stanford, P. K., & Kitcher, P. 1997. "Refining the Causal Theory of Reference for Natural Kind Terms." *Philosophical Studies* 97:99–129.