Facta Philosophica 1, 1999: 135-165

Moral Realism and Twin Earth

Stephen Laurence, Eric Margolis & Angus Dawson

Hilary Putnam's Twin Earth thought experiment has come to have an enormous impact on contemporary philosophical thought. But while most of the discussion has taken place within the context of the philosophy of mind and language, Terence Horgan and Mark Timmons (H&T) have defended the intriguing suggestion that a variation on the original thought experiment has important consequences for ethics.[1] In a series of papers, they've developed the idea of a *Moral Twin Earth* and have argued that its significance is that it has the resources to undermine naturalistic versions of moral realism.[2] H&T don't hold back in their assessment. "Moral Twin Earth", they say, "packs a mean punch", and ethical naturalism is "down for the count" (H&T 1990/91, p. 461). "[I]n the end, all defensive strategies are likely to prove futile against Moral Twin Earth" (H&T 1992b, p. 171). H&T aren't the only ones who think this. R. M. Hare endorses H&T's strategy, describing their case against ethical naturalism as "effective" and "illuminating" (Hare 1995, p. 342 & p. 352).

Unfortunately, H&T's use of Moral Twin Earth resists a quick summary. This is because the thought experiment plays into not one, but three distinct arguments. The first two require a good amount of stage-setting and are supposed to revive classic arguments against ethical naturalism–J. L. Mackie's argument from queerness and G. E. Moore's open question argument. The third, which we call the direct argument, is less explicit in H&T's writings, but it is doing at least as much work for H&T as the other two.

The three arguments reinforce one another, so H&T's case against ethical naturalism may look daunting. However, appearances are deceptive. We will argue that, in the end, H&T's arguments provide no reason at all for rejecting ethical naturalism. Moral Twin Earth neither revives the classic arguments

1 See H&T (1990/91), (1992a), (1992b).

2 For some recent versions of naturalistic moral realism see Boyd (1988), Brink (1984, 1989), and Railton (1986).

against ethical naturalism nor does it undermine ethical naturalism on its own.

Here's how we proceed. In section 1, we lay out the basic thought experiment, showing how Moral Twin Earth is supposed to be a variation on Putnam's original thought experiment. In section 2, we take up H&T's version of the argument from queerness. In its revised form, the argument is supposed to show that, under naturalistic assumptions, ethical properties are unacceptable because their supervenience on physical properties cannot be explained. In section 3, we turn to H&T's version of the open question argument. In its revised form, the argument is supposed to show that there is an important asymmetry between paradigmatic a posteriori identity claims such as water = H_2O and claims that offer corresponding identities between moral properties and natural ones. The former, but not the latter, can be established by reflecting in a prescribed way on our semantic competence with the terms involved. Moral Twin Earth figures in these first two arguments by supporting crucial premises. But it isn't until we get to the direct argument, in section 4, that we can see how powerful the thought experiment is supposed to be. At this point H&T argue that it simply follows from the thought experiment that moral terms can't be rigid designators and that ethical naturalism is flawed for this reason alone. Again, we'll argue that, despite H&T's persistent efforts to undermine ethical naturalism, not one of their arguments is successful. Ethical naturalism may have its problems, but Moral Twin Earth is not among them.

1 From Twin Earth to Moral Twin Earth

Though later (in sec. 4.1) we'll argue that H&T's characterization of Moral Twin Earth is deeply misleading, in this section we'll hold off on criticism and introduce their thought experiment as they themselves do. Since Moral Twin Earth is a variation on the standard Twin Earth scenario, it helps to begin with Putnam's original thought experiment.

In the original thought experiment (Putnam 1973, 1975), we are to imagine that there is a place that is virtually identical to Earth except that, where H_2O fills our lakes and streams (and so on), a liquid that is perceptually indistinguishable from H_2O, but has a different chemical composition—"XYZ"—fills theirs. Apart from this one difference, Twin Earth is supposed to be exactly the same as Earth so that Twin Earth even has doppelgangers corresponding to everyone on Earth.

Twin Earth is a philosophical fantasy that's supposed to bear upon the semantics of natural language. The value of the thought experiment is that it isolates several factors that ought to be disentangled in a full account of

meaning, especially the contribution of one's beliefs about a kind and the contribution of one's environment. To see what's at stake, consider the obvious fact that in using a word like "water", "gold" or "cat", one has encountered (at most) a limited number of samples or instances to which these words apply. Yet it's patent that these words apply to other, unencountered items. The question, then, is which other things fall within the application of these words. What Putnam's Twin Earth thought experiment allows one to do is examine this question while paying careful attention to the issue of how a speaker's beliefs and her environment affect the answer.

In one version of the thought experiment, Putnam asks us to go back to a time when people on Earth lacked a sufficiently developed chemistry so that they knew nothing about the chemical composition of the liquid that filled their lakes, streams, and so on; no one had even heard of H_2O. All the same, Putnam has the intuition–and urges us to have the intuition–that the word "water", at that time, applied to H_2O only and not to XYZ despite the fact that no one was in a position to discriminate between the two. The conclusion he draws from this sort of case is that meaning for natural kinds is determined to an important extent by the environment. What makes something fall under "water" is that it bears the *same liquid* relation to certain paradigmatic samples. More generally, the claim is that natural kind terms apply to all and only those things that have the same essence as the paradigmatic exemplars within a linguistic community and consequently that the meaning of a natural kind term is partly determined by facts about the environment.[3]

Sometimes the key intuition that underwrites Putnam's conclusion is elicited by the question of whether an Earthling and his twin mean the same thing when they utter the same form of words. The point of putting things this way isn't just to compare the semantics of two languages (English and Twin-English), but to highlight the importance of the environment to the meanings of our own linguistic expressions. The standard intuition in such cases is that, when an Earthling says "...water..." and his twin says "...water...", they mean different things; the Earthling's word "water" refers to H_2O, while his twin's word "water" refers to XYZ. And it is by recognizing this difference, that one is supposed to be able to see that the meaning of *our* word "water" depends on external facts, that is, facts about the nature of the environment in which we, English speakers, inhabit.

Now generally speaking the kinds of examples used in discussions of Twin Earth are ones where the essence of a kind is identified with some

3 This is the inspiration behind externalist theories of content, aptly summarized by Putnam's slogan that "'meanings' just ain't in the *head*!".

aspect of its internal structure–water, gold, aluminum. Yet in principle the same considerations hold in cases where the essence of a kind is its having a particular functional property. For example, supposing that the essence of a biological organ is its typical causal role within a larger biological system, one can imagine that a species of Twin animals have a biological organ that is superficially like its counterpart on Earth but one that exhibits a distinct causal role despite appearances to the contrary.[4]

The Moral Twin Earth thought experiment capitalizes on this type of possibility, where two properties or kinds have distinct functional roles yet share many of their more salient superficial aspects. The variety of ethical naturalism that H&T take as their principal target is one where ethical properties are supposed to be functional properties whose causal roles are articulated by a normative moral theory (see, e.g., Brink 1984). What we are to imagine is that Moral Twin Earth is just like Earth except in one crucial respect: It differs with respect to the relevant functional properties. As an illustration, H&T ask us to suppose that on Earth "moral judgments and moral statements are causally regulated by some unique family of functional properties whose essence is functionally characterizable by the generalizations of some [particular] ... consequentialist theory, which we will designate T^c" (H&T 1992a, p. 245). On Moral Twin Earth, "the properties tracked by Twin English moral terms are also functional properties, whose essence is functionally characterizable by means of a normative moral theory. But these are *non-consequentialist* moral properties, whose functional essence is captured by some specific deontological theory; call this theory T^d" (H&T 1992a, p. 245).

In stating the details of the thought experiment here, H&T appeal to Richard Boyd's causal theory of reference (*causal regulation* is shorthand for satisfying the conditions of Boyd's theory).[5] Be this as it may, the exact nature of Boyd's theory isn't supposed to be important, nor is it supposed

4 The biological example helps to illustrate the way that functional essences and their superficial indices can be teased apart, but this doesn't mean that we are endorsing a non-historical treatment of the individuation of biological kinds.

5 In summarizing the theory, H&T quote from Boyd (1988):

> *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* .... 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* ..., a pattern of deference to experts on *k* with respect to the use of *t*, etc... (quoted in H&T 1992b, pp. 158–9).

to matter exactly which functional properties are assumed to be on Earth and on Twin Earth. The crucial part of the thought experiment is just that, whatever sort of functional properties get tracked by moral terms on Earth—a matter which, by hypothesis, is open to empirical investigation—different functional properties are tracked on Moral Twin Earth.

Having laid out the thought experiment, H&T ask whether the intuitions that it prompts are analogous to the intuitions that are standardly associated with Twin Earth. Their claim is that they are not. In particular, H&T maintain that one doesn't have the intuition that English terms such as "right" and "wrong" refer to different properties than their Twin English counterparts. Instead, "the natural response to the differences contemplated in the Moral Twin Earth Story is that Earthlings and Twin Earthlings differ in their respective moral *beliefs*, and ultimately differ in the respective moral *theories*" (1992a, pp. 247–8).[6] Recall that with Putnam's original example, people don't want to say merely that Twin Earthlings have different beliefs about water. It's this asymmetry that makes Moral Twin Earth such an important tool for H&T. The asymmetry feeds into three arguments against ethical naturalism. We are now in a position to turn to these arguments. We'll begin, in the next section, with H&T's treatment of the argument from queerness.

2 The New Argument from Queerness

Mackie's original argument from queerness expresses the concern that the properties to which the moral realist is committed are just too strange or "queer" to fit into a naturalistic picture of the world and that we should therefore not suppose that they are real.[7] One way of construing the demands of naturalism—a way that is reinforced by examples like water/H_2O—is in terms of the identification of questionable properties with natural or descriptive properties. Then the charge might be that ethical properties are queer in the sense that they are not identical to any natural or descriptive properties.[8]

However, as H&T point out, the ethical naturalist might respond that one needn't advocate such a strong thesis as this to be an ethical naturalist.

6 See also H&T (1992b), pp. 165–6; (1990/91), p. 460.

7 Mackie's discussion is directed towards what he calls the *objectivist*, by which he seems to mean an advocate of a non-naturalistic form of moral realism. He seems to think that any sensible naturalist wouldn't advocate the existence of moral properties. For his argument, see Mackie (1977), pp. 38–43.

8 The notions of "natural" and "descriptive" properties (which may not be equivalent) are notoriously slippery. At a minimum, the properties appealed to in physics and other natural sciences qualify, and ethical properties, at least *prima facie*, do not.

As H&T note, one of the central themes of recent philosophy of mind has been the rejection of a strong reductive physicalism (i.e., the kind that requires property identities) on the grounds that mental properties are multiply realizable. The multiple realizability of the mental has suggested to many that a weaker understanding of physicalism is required. A popular way of explicating this weaker sense of physicalism has been in terms of the notion of supervenience. It's against the background of these ideas that H&T develop their own version of the queerness argument.

Supervenience, as David Lewis says, is the idea that "there could be no difference of one sort without differences of another sort" (Lewis 1986, p. 14).[9] For example, temperature supervenes on mean kinetic energy, since there could be no difference in temperature without a difference in mean kinetic energy.[10] To a first approximation, a supervenience-based physicalism holds that properties are acceptable only when they supervene on the physical. On the face of it, however, a constraint of this kind is perfectly congenial to ethical naturalism. As H&T themselves note, people generally agree that moral properties do supervene on natural or descriptive properties. In fact, the concept of supervenience, as it's understood today, emerged in ethical theory.[11] So given a supervenience-based physicalism, moral properties appear to be relatively unproblematic.

These facts about supervenience provide the starting point for H&T's new argument from queerness. They grant that an appeal to supervenience provides the ethical naturalist with a little breathing room, but then they argue that ultimately this appeal won't work. Though their argument gets quite intricate, the main idea is this. They claim that supervenience relations can't be accepted unless they themselves can be explained; unexplained supervenience relations are supposed to be "queer"—that's the connection with Mackie's original argument. H&T then go on to prescribe a specific explanatory strategy, which they illustrate with the case of mental properties. The problem for the ethical naturalist is supposed to be that the same prescribed explanatory strategy doesn't work for ethical properties. Though ethical properties supervene on the physical, H&T maintain that

9 Supervenience has been widely discussed recently. A number of different varieties of supervenience have been distinguished, and their properties and relations have been investigated in some detail (for discussion, see McLaughlin 1995 and Kim 1984). For our purposes, however, Lewis's characterization of the basic idea is all we need.

10 For the sake of the example, we ignore some of the complexity of the physics of temperature in different media.

11 The notion of supervenience in roughly its modern form is often traced to Hare (1952). See Kim (1984) and McLaughlin (1995) for further discussion of the history of the concept of supervenience.

there is no explanation for why they do and that this suffices to discredit them.[12]

H&T begin by claiming that "supervenience in ethics is, at a minimum, a *semantic* constraint upon moral language and moral judgment" (H&T 1992a, p. 233; emphasis added). This strikes us as a strange move. Why abandon the standard characterization of supervenience in terms of an *ontological* relation among classes of properties? In the end, we don't see that they have any motivation for switching the orientation to language, and it's their failure to take seriously the ontological characterization of supervenience that undermines their new argument from queerness.[13] We'll put these worries to the side for the moment though and concentrate on H&T's standard for explaining supervenience relations.

If some higher level property P supervenes on the physical, this means that there can be no difference with respect to P without a difference with respect to the physical. So it isn't possible to have the world be physically exactly the same as it is now and yet differ with respect to P. And, in general, there won't be any two possible worlds—any two complete ways that the world might be—that are the same physically but differ with respect to P. The question, then, is why this should be. H&T write (H&T 1992a, p. 234),

> From a broadly naturalistic perspective, it seems that the appropriate answer is just this: given two sufficiently detailed partial descriptions D1 and D2 of two such putative worlds, together perhaps with a sufficiently detailed partial description D of our actual world, at least one of the descriptions D1 and D2 will contain violations of the semantics of certain terms and concepts.

H&T go on to develop their schema for explaining supervenience by isolating two types of semantic constraints. The first, which they call *pure semantic constraints,* concern the proper use of a term insofar as it involves semantic knowledge that must be mastered by anyone who is able to use it correctly. As an example, H&T cite the constraint given by (P1), which embodies the view of natural kind terms that Kripke and Putnam are famous for having argued for (H&T 1992a, p. 235):

> (P1) For any physical-stuff natural kind term t, and any physico-chemical property P, if S refers at our actual world to physical stuff with a distinctive physical essence, and this stuff's physical essence is its possession of P, then t refers rigidly (i.e., at every possible world) to stuff that possesses P.

12 Though we won't pursue this line of response, it is possible to question whether it's really necessary that such supervenience relations be explained in order for the properties in question to be legitimated. Here we just note that H&T don't provide any support for the claim that it is necessary.

13 We take up both of these points towards the end of this section.

Second, H&T mention the existence of what they call hybrid semantic constraints. These incorporate semantically relevant empirical facts. For instance, given that "water" actually refers to stuff "whose distinctive physical essence is to be composed of H_2O molecules", we have the following hybrid constraint based on (P1) (H&T 1992a, p. 235):

> (H1) For any possible world w, a quantity of liquid (in w) belongs to the w-extension of 'water' iff it is composed of H_2O molecules (in w).

Let's now turn to H&T's proposed explanation of the supervenience of the mental on the physical. Again, it's important to see how this case plays out since it's here that they establish a prescribed general method for explaining supervenience relations. Their explanation for psychological properties is based on psychofunctionalist accounts of the propositional attitudes. According to psychofunctionalism, propositional attitude state types (states like believing that it may rain today or hoping that it won't) are essentially characterized by their causal role, as given by a set of psychological laws. A complete psychological theory would describe this set of laws, thereby specifying a functional role for each propositional attitude state type. This generates the following pure semantic constraint (1992a, p. 237):[14]

> (P2) For any psychological theory T, if (i) there is some unique family of interconnected functional properties that causally regulate (in the actual world) the attributions by humans of propositional attitudes to one another and to themselves, and (ii) the generalizations of T collectively characterize the functional essence of these properties, then each propositional-attitude term rigidly refers to the T-characterizable functional property that regulates it.

Suppose now that we take some particular psychological theory T*, and T* posits the existence of propositional attitudes and, at the same time, satisfies the conditions of (P2). These are empirical facts which, together with (P2), yield the following hybrid semantic constraint (H&T 1992a, p. 238):

> (H2) For any possible world w, the correct assignment of w-extensions to propositional attitude terms is an assignment which renders true (at w) all the generalizations of T*.

On the basis of (P2) and (H2), H&T provide the following as "the appropriate general form of explanation" for the supervenience of a given propositional attitude type, M, on the physical (H&T 1992a, p. 238):

14 H&T say that psychofunctionalism actually asserts this constraint. However, psychofunctionalism is standardly taken to be a theory of the nature of psychological states, not a theory about the referential properties of psychological terms.

> P is a member of a system of physiochemical properties which together more or less realize the pattern of causal relations characterized by the psychological generalizations which semantically constrain mental terms and concepts. Furthermore, P itself occupies the role in that system which, according to those psychological generalizations, constitutes the M-role. Hence, whenever someone instantiates P, he must instantiate M as well.

In other words, there should be no doubt about why the propositional attitudes supervene on the physical. Given the semantic rules governing the use of mental terms and concepts and the crucial empirical facts that H&T assume for purposes of argument, instantiating P guarantees the instantiation of M.

With an explanation of why the mental supervenes on the physical, H&T turn to the question of whether the same strategy works with the ethical. Recall that the variety of ethical naturalism at stake corresponds to the psychofunctionalist treatment of mental properties. Ethical properties, that is, are taken to be functional properties whose causal roles are specified by a normative moral theory. With this view as their target, H&T ask what semantic constraints would be implicated in an explanation of the super-venience of the ethical on the physical, an explanation that's supposed to be analogous to their explanation of the supervenience of the mental on the physical. First we have the pure semantic constraint (P3) (H&T 1992a, p. 243):

> (P3) For any normative moral theory T, if (i) there is some unique family of interconnected functional properties that causally regulates the actual-world moral judgments and moral statements of humans, and (ii) the generalizations of T collectively characterize the functional essence of these properties, then each moral term refers rigidly to the T-characterizable functional property that regulates it.

Then, (P3) together with some empirical facts (including the fact that the particular theory T^* is a complete and correct normative moral theory) generates the following hybrid semantic constraint (H&T 1992a, p. 243):

> (H3) For any possible world w, a correct assignment of w-extensions to moral terms is an assignment which renders true (at w) all the generalizations of T^*.

We can now see how H&T's argument is supposed to work. By their lights, if moral terms don't conform to (P3) and (H3), then the supervenience of the ethical on the physical remains unexplained, leaving moral properties outside our otherwise naturalistic conception of the world. For H&T, this is what saying ethical properties are "queer" amounts to. Moral Twin Earth enters the picture because it is supposed to undermine (P3) and (H3) by showing that moral terms don't rigidly designate moral properties, that is,

they don't refer to the same properties in all possible worlds where they have a referent.

Recall that in the Moral Twin Earth thought experiment different functional properties are taken to causally regulate the moral terms on Earth and Twin Earth; properties that are specified by a consequentialist theory are tracked in the one case, and properties that are specified by a deontological theory are tracked in the other. The intuition that H&T encourage is that people on Earth and their doppelgangers on Twin Earth have different beliefs *about the very same ethical properties*—a striking contrast to the standard reading of the H_2O/XYZ case. Suppose we grant the intuition.[15] The consequence, if we take it at face value, is that moral terms don't rigidly designate moral properties. Since embedded in (P3) is the claim that moral terms do rigidly designate moral properties, (P3) can't be right, and, in particular, can't be used in an explanation of the supervenience of the ethical on the physical. As H&T summarize the situation, "The immediate upshot of the Moral Twin Earth thought experiment is that synthetic moral functionalism is not tenable and, hence, cannot undergird SCS explanations [i.e., semantic constraint satisfaction explanations] of putative objective supervenience relations between nonmoral and moral facts or properties" (H&T 1992a, p. 248).

Now this argument is flawed in a number of respects, but the fundamental problem is H&T's tendency to think of supervenience as a semantic relation, rather than the usual way, where supervenience is taken to be a metaphysical relation. It turns out that when the task of explaining a supervenience relation is construed as a task of explaining a metaphysical relation, there are alternatives to the strategy that H&T recommend for explaining such relations. So even if the sort of semantic explanation they are interested in isn't applicable in the ethical case, the supervenience of the ethical on the physical *can* be explained. Moreover, there doesn't seem to be any good motivation for pursuing a H&T-style semantic reading of supervenience.

To begin, it's the hybrid constraints that do most of the work for H&T. They are "hybrids" in that they are supposed to mix semantic information with empirical fact. But it is a trivial matter to generate a purely metaphysical version of any of these constraints. Consider H&T's original constraint, (H1).

> (H1) For any possible world w, a quantity of liquid (in w) belongs to the w-extension of "water" iff it is composed of H_2O molecules (in w).

Here's a metaphysical version:

> (H1*) For any possible world w, a quantity of liquid (in w) is water iff it is composed of H_2O molecules (in w).

15 In section 4 we argue that the intuition shouldn't be granted.

Or, turning to their relevant point of comparison, consider H&T's hybrid constraints for propositional attitudes and ethical properties:

> (H2) For any possible world w, the correct assignment of w-extensions to propositional attitude terms is an assignment which renders true (at w) all the generalizations of T*.
>
> (H3) For any possible world w, a correct assignment of w-extensions to moral terms is an assignment which renders true (at w) all the generalizations of T*.

These too have straightforward metaphysical versions:

> (H2*) For any possible world w, the things that are propositional attitudes in w are those things which render true (at w) all the generalizations of T*.
>
> (H3*) For any possible world w, the things that are moral properties in w are those properties which render true (at w) all the generalizations of T*.

Given the availability of these metaphysical correlates to H&T's constraints, it's only a short step to giving a metaphysical analog to their explanation of supervenience as well. For example,

> P is a member of a system of physicochemical properties which together more or less realize the pattern of causal relations which characterize the system of propositional attitude state types. Furthermore, P itself occupies the role in that system which constitutes some particular propositional attitude, M. Hence, whenever someone instantiates P, he must instantiate M as well.

But once we've gone this far, it is not the least bit difficult to provide exactly the same type of explanation for why the ethical supervenes on the physical:

> P is a member of a system of physicochemical properties which together more or less realize the pattern of relations which characterize the system of ethical state types. Furthermore, P itself occupies the role in that system which constitutes some particular moral property, M. Hence, whenever someone instantiates P, he must instantiate M as well.

What's more, since the Moral Twin Earth thought experiment is aimed at the semantic element of H&T's semantic constraints, it fails to undercut this metaphysical analog to their explanation.

We get entirely parallel explanations of the supervenience of the psychological and the ethical on the physical, once the gratuitous detour through

semantics is omitted. So it looks like the challenge from H&T's revised argument from queerness is easily met.

In sum, the problem with H&T's revised argument from queerness is that once the inessential semantic baggage is removed, it doesn't provide any reason for thinking that ethical properties are worse off than psychological ones. To the extent that it's necessary to explain supervenience relations at all, the natural explanation to seek is a metaphysical one since supervenience itself, as it is standardly conceived, is a metaphysical relation. And once H&T's explanations are recast in metaphysical terms, the result is entirely parallel explanations of the supervenience of the psychological and the ethical on the physical. So it looks like the challenge from H&T's revised argument from queerness is easily met.

Still there are some residual worries that we should briefly comment on before turning to H&T's next argument. First, why do H&T opt for the strategy of providing *semantic* explanations of supervenience relations? And second, is there really an asymmetry between the ethical case and the psychological case for semantic explanations of the sort that H&T favor? We'll take up these questions in reverse order.

The asymmetry between the psychological and the ethical case vis-a-vis semantic explanations of supervenience would certainly be puzzling if it were real. But it still wouldn't support H&T's revised argument from queerness, since that argument turns on the claim that there is *no* explanation of the supervenience of the ethical on the physical. Since we've just seen there is an explanation of this relation—one which closely mirrors the sort of explanation H&T themselves want to give—the queerness argument is defeated. Perhaps this asymmetry could be parlayed into a separate argument against ethical naturalism, but that remains to be seen. In any case, in section 4 we will argue that the asymmetry is illusory: when the Moral Twin Earth thought experiment is described more carefully, it vanishes completely.

But why do H&T insist on a semantic explanation of supervenience in the first place? It's hard to say. Since they don't explain why they adopt this strategy, we can only guess at their reasons. As best as we can tell there are two possibilities, neither of which holds up to scrutiny. The first is to allow some room for noncognitivists to maintain a supervenience thesis about the ethical. Indeed, H&T remark that most philosophers agree that the ethical supervenes on the physical, including non-cognitivists. But since non-cognitivists don't believe in moral properties, facts, and relations, they have to frame their commitment to supervenience as a point about the "logic of moral discourse". H&T cite Hare (1952) and Blackburn (1984) as giving examples of unproblematic ways in which noncognitivists might explain

supervenience. So perhaps H&T opt for a semantic construal of supervenience in order to cover all metaethical theories at once.

If this is their line of thinking, however, it isn't convincing. As H&T themselves point out, noncognitivists are in a very different position than ethical naturalists. Since noncognitivists don't believe in moral facts, they are subject to "no burden of explaining such facts" (H&T 1992a, p. 231). Maybe they do have to explain certain peculiar features of moral discourse, but why think that what needs to be explained here is related in any interesting way to what needs to be explained for someone who thinks that there are ethical facts?[16] And since the explanada aren't at all alike, why try to frame the two explanatory burdens in general enough terms to cover both? As we see it, noncognitivism offers little motivation for requiring a semantic construal of supervenience in the general case.

A second reason that H&T might have for a semantic reading of supervenience is to accommodate nominalists (see H&T 1992a, p. 235). Since nominalists don't accept any properties into their ontology, they can't accept moral properties. Still, they may want to defend a kind of supervenience thesis about the ethical. The way to do this is via semantic assent. That is, rather than saying that such-and-such class of properties (e.g., temperature) supervenes on so-and-so other class of properties (e.g., mean kinetic energy), one could claim that truths involving such-and-such a class of *predicates* (temperature predicates) supervene on truths about so-and so other class of *predicates* (mean kinetic energy predicates).

In this case, however, it would be rather odd to describe the situation as one where supervenience offers a semantic constraint or that supervenience is ultimately a semantic phenomenon. This is for the simple reason that the relations among the truths that the nominalist posits hold (in the standard case) because of the way the world is. For example, if a nominalist were to maintain that temperature supervenes on mean kinetic energy, that's because she believes that there couldn't be a change in temperature without a change in mean kinetic energy, not because she believes that there is some semantic constraint on our concepts of temperature and mean kinetic energy. The problem with the claim that "two objects with exactly the same mean kinetic energy have different temperatures" isn't that it's semantically

16 Actually, we doubt that that the explananda is semantic even for noncognitivists. There is certainly *something* wrong with someone who calls one act right but isn't willing to call another act right even though the second has all of the same non-moral features as the first. Yet just because there is something wrong with such a person doesn't mean that she has violated a linguistic rule. Indeed it's hard to see why an emotivist, e.g., would think that the situation involves anything more than an inconsistency in emotional reaction.

deviant; the problem is that it's *false*. So nominalism doesn't provide any motivation for H&T's semantic construal of supervenience either.

We aren't sure which of these considerations is driving H&T's detour through semantics. In the end, however, we don't think it matters. Neither works, and neither really plays a substantial role in their discussion. Having apparently used one or the other or both of them to motivate treating supervenience as a semantic constraint, H&T proceed to ignore noncognitivists and nominalists alike; the discussion focuses wholly on ethical naturalism. In their pure semantic rules—e.g., (P1)—they explicitly refer to properties, abandoning the concern with nominalism, and they even remark, at one point, that their "operative constraints need not necessarily be regarded as primarily a matter of language per se." They go on to say that "[u]nder a realist position concerning universals, the operative constraints ultimately describe what is constitutive of the mind-independent attributes and kinds predicated by various terms of our language; the constraints apply to terms only derivatively, by virtue of applying to the attributes and thing-kinds those terms pick out" (H&T 1992a, p. 235). This, of course, is a major concession. Since the target of their critique is a theory that actually postulates the existence of properties—moral properties—then, by their own lights, H&T have no reason to assess the theory in terms of their semantically-couched constraints. The straightforwardly metaphysical alternatives that we've spelled out should do. And since these don't pose any special difficulty for ethical naturalism, H&T aren't in a position to say that ethical properties are queer.

In short, Moral Twin Earth doesn't help in the least to revive Mackie's argument from queerness. Still, H&T have two other arguments to call upon. Let's turn to their revised version of Moore's open question argument.

3 The New Open Question Argument

Moore (1903) assumes that any reductive treatment of the ethical would reflect analytic truths and, as a consequence, be knowable a priori. The problem, according to Moore, is that for any proposed reduction (e.g., good is N, where N is some naturalistic property), one can always entertain the question whether the reduction is correct (e.g., one can wonder whether something that is N is good). For Moore, this is to say that the question is open. Under the assumption that a reduction is knowable a priori, however, there shouldn't be any question of whether the reduction is correct; mere reflection should suffice to show that it is. Moore's conclusion is that no reductive account of the ethical can stand up to scrutiny.

The natural response to Moore is, of course, that reductions needn't be analytic (and hence, needn't be a priori). Nowadays the point is often made

by appealing to Kripke's (1972) and Putnam's (1970, 1973, 1975) work on the causal theory of reference and the sharp distinction between metaphysics and epistemology that goes along with that work. Following Kripke and Putnam, one might hold that a reductive account of goodness only requires a synthetic property identification. What's more, if we take scientific accounts of natural kinds to provide property identifications—as Kripke and Putnam do—then it's patent that property identifications needn't be analytic. Certainly, "water is H_2O" and "heat is molecular motion" aren't analytic. So why ask anything more of ethical properties? In fact, the apparatus that Kripke and Putnam introduced isn't really necessary to make the basic point here. All that's needed is the idea that co-referential terms can be non-synonymous, a point that goes back at least to Frege's (1892) distinction between sense and reference. So it's puzzling that this moral should have taken so long to be drawn and that Moore's argument should have been so influential in the first place.

Be that as it may, the argument certainly has been influential, and H&T believe that it can be revived and put to work against recent versions of ethical naturalism that make no attempt at analytic reductive accounts of ethical kinds. Once again, at the center of their argument is Moral Twin Earth. H&T's strategy is to show that Moral Twin Earth demonstrates that certain key questions concerning moral terms are left open, where the corresponding questions are closed for terms like "water". They provide the following updated account of what it is for a question to be open or closed (1992b, p. 161):

> [L]et us say that a question is *closed* just in case most any semantically competent speaker who considers the question carefully, and who properly brings his semantic competence to bear on the question, will judge both that the answer to the question is obviously 'yes' (or obviously 'no'). The idea is that semantic competence alone, apart from any specific empirical knowledge the speaker might possess, is the likely source of the judgment; and that the intuitive obviousness of the answer is evidence that this is its source. Let us say that a question is open just in case it is not closed.

In addition, they note that "if a question has any empirical assumptions, then knowing the answer amounts to knowing what answer would be correct if the assumptions were all true" (1992b, p. 171).

Moore's open questions for the case of goodness are Q1 and Q2.

Q1. Act A is N, but is it good?

Q2. Act A is good, but is it N?

The direct analogs to these questions in the water/H_2O case are Q3 and Q4.

Q3. Liquid L is H_2O, but is it water?

Q4. Liquid L is water, but is it H_2O?

To update these for the revised version, H&T suggest replacing Q3 and Q4 with Q5 and Q6. These "have built into them the appropriate empirical hypothesis about causal regulation"; that is, they incorporate Richard Boyd's theory of reference, which H&T adopt for purposes of argument (H&T 1992b, p. 162).

Q5. Given that the use of "water" by humans is causally regulated by the natural kind H_2O, is liquid L, which is H_2O, water?

Q6. Given that the use of "water" by humans is causally regulated by the natural kind H_2O, is liquid L, which is water, H_2O?

H&T claim that most any competent speaker of English will find the answer to these questions to be obvious. Hence, by their criteria, these questions are closed. In contrast, the analogous questions for ethical properties, Q7 and Q8, are supposed to be open.

Q7. Given that the use of "good" by humans is causally regulated by natural property N, is act A, which has N, good?

Q8. Given that the use of "good" by humans is causally regulated by natural property N, does act A, which is good, have N?

Their reasoning is that Moral Twin Earth shows that our semantic intuitions fail to generate an obvious answer to these questions.

Unfortunately, one of the difficulties in assessing this argument is that H&T fail to develop it. They never come out and say explicitly how Moral Twin Earth bears upon the issue of whether these questions can be answered merely by consulting our semantic intuitions.[17] Nevertheless, we think it's fairly clear that they couldn't have much of an argument because of certain assumptions they make about the questions concerning non-moral properties and kinds, that is, Q5 and Q6. However their argument is supposed to go, it is going to depend crucially on the claim that, in bringing one's semantic competence to bear upon Q5 and Q6, one will inevitably find the answers to be obvious. This being the case, the questions are supposed to be closed (i.e., according to H&T's definitions). But is it really the case that the answers are obvious in this way? The answer is most certainly that they are not. And the reason they are not is because, even if Boyd's theory of reference is true, it's not obviously true and, more to the point, it's not analytic. At

17 The reader should consult H&T (1992b), pp. 166–7. Here H&T state that the Moral Twin Earth thought experiment establishes that Q7 and Q8 are open, but never say why. See also H&T (1990/1991) p. 461, which leaves the argument equally undeveloped.

most, Boyd's theory is an a posteriori synthetic truth. Given this, adding that this theory is satisfied can't turn an open question into a question whose answer is obvious simply by virtue of one's semantic competence.

This point is easy to miss. What H&T build into Q5 and Q6 is the empirical information that Boyd's theory is satisfied for the relevant terms—that is, that "water" is causally regulated by H_2O. But surely this is not enough to close these questions for speakers who are supposed to be consulting their semantic knowledge. The problem is that these speakers have no way of knowing whether Boyd's causal regulation theory is the true theory of reference. After all, causal theorists like Boyd and Putnam do not claim that such theories are analytically true—that their truth is known on the basis of a speaker's semantic competence alone.

It's easy to miss this point because the answers to Q5 and Q6 *do* seem obvious to ordinary English speakers. Yet this is simply because it is such a familiar everyday fact that water is H_2O. The familiarity of this fact makes it all too easy to ignore the whole point of Q5 and Q6, which is to determine the answers to these questions based on "semantic competence alone, apart from any specific empirical knowledge the speaker might possess" (1992b, p. 161). "Obviously, yes" one might say, simply on the grounds that everyone knows that water is H_2O. But Q5 and Q6 aren't about whether these are facts that we all know to be true. The thing that matters is whether we can know their answers based on semantic competence alone.

It may help matters to take a less well-known identity statement from chemistry. So Q5 and Q6 might be replaced by something like:

> Q9. Given that the use of "saccharin" by humans is causally regulated by $C_7H_5O_3NS$, is substance S, which is $C_7H_5O_3NS$, saccharin?
>
> Q10. Given that the use of "saccharin" by humans is causally regulated by $C_7H_5O_3NS$, is substance S, which is saccharin, $C_7H_5O_3NS$?

Since people generally know nothing about the chemical structure of saccharin, if these questions were obvious, it wouldn't be because of any familiarity with $C_7H_5O_3NS$. But it seems clear that Q9 and Q10 aren't obvious. At a minimum, speakers will need to know whether causal regulation guarantees reference—and this, as we have just noted, is not something that can be determined soley on the basis of a speaker's semantic competence. So Q5 and Q6 end up being open just as much as Q7 and Q8 once serious due is given to the conditions that H&T specify for determining whether a question is open or closed.

Perhaps we could just build into Q5 and Q6 the information that we are claiming speakers require to settle these questions, following H&T's

qualification that "if a question has any empirical assumptions, then knowing the answer amounts to knowing what answer would be correct if the assumptions were all true". That is, perhaps we could just build in the fact that Boyd's causal regulation theory of reference is true. The trouble is, this information doesn't seem to be "an empirical assumption" for Q5 and Q6; adding this information simply changes the questions being asked. Still, if the information were added, we'd end up with two new questions along the lines of Q11 and Q12:

> Q11. Given (i) that the use of "water" by humans is causally regulated by the natural kind H_2O and (ii) that the causal regulation theory of reference is the true theory of reference, is liquid L, which is H_2O, water?
>
> Q12. Given (i) that the use of "water" by humans is causally regulated by the natural kind H_2O and (ii) that the causal regulation theory of reference is the true theory of reference, is liquid L, which is water, H_2O?

But if we are going to do this, we might as well skip mentioning Boyd's particular theory of reference altogether,[18] and just say that the terms refer to the kinds. This, however, generates, questions Q5'–Q8'.

> Q5'. Given that "water" refers to the natural kind H_2O, is liquid L, which is H_2O, water?
>
> Q6'. Given that "water" refers to the natural kind H_2O, is liquid L, which is water, H_2O?
>
> Q7'. Given that "good" refers to the natural property N, is act A, which has N, good?
>
> Q8'. Given that "good" refers to the natural property N, does act A, which is good, have N?

Now Q5' and Q6' are in fact closed. But the trouble is that Q7' and Q8' are also closed. So once again the question are all on a par. This time they are all closed. Having been told that a word refers to some property or kind, there is nothing left to wonder about. The question can't help but be closed because it's been framed in a way that guarantees assent. The bottom line is that, in order to formulate the water/H_2O question in a way that keeps it

18 Why say that "X" is causally regulated by X and that causal regulation is the true theory of reference when you can simply say that "X" refers to X?

Note as well that nothing in our discussion turns on the specifics of Boyd's theory of reference, since the only fact about Boyd's theory that we appeal to is that it is not analytically true. Since the synthetic and a posteriori character of causal theories of reference is not in dispute, any other theory would have the same consequences.

closed by H&T's standards, the parallel question about goodness is going to be closed as well.

Once the questions are properly framed, the alleged asymmetry between ethical kinds and other natural kinds vanishes. In particular, once we insist that our answers derive purely from our semantic competence (as H&T would have it), then the only way to close the questions about non-ethical kinds is to stipulate that the terms in question refer to the relevant kinds. But following this strategy means that the parallel questions about ethical kinds will be closed as well. Once again, then, there's little to be said for H&T's claim that they've managed to revive a classic argument against ethical naturalism.

4 A Direct Argument against Ethical Naturalism

While H&T bring considerable attention to the argument from queerness and the open question argument, their fondness for Moral Twin Earth goes deeper than that. Often they sound as if they think that Moral Twin Earth by itself suffices to undermine ethical naturalism.[19] The consideration that appears to move them is that Moral Twin Earth shows, in their view, that moral terms fail to designate the same physical or functional property in every possible world where they have a referent; that is, they think that Moral Twin Earth shows that moral terms aren't rigid designators. In contrast, "water", "gold", and comparable natural kind terms do designate the same property in every possible world where they have a referent. Perhaps, then, what really underlies H&T's rejection of ethical naturalism is this purported difference. The thought must be that, if ethical terms aren't rigid designators, that's because there aren't any moral properties in the first place or because moral terms designate different properties in different worlds. Either way the non-relativist, ethical naturalist loses.

Still, she only loses if H&T are right in concluding that ethical terms aren't rigid designators. In this section, we will examine H&T's argument that they aren't. What's at stake for H&T, however, is more than a single argument against ethical naturalism. On the contrary, if the present argument fails to work, so do H&T's other two. Both crucially rely on the purported result that moral terms aren't rigid designators.[20] If H&T can't show

19 Indeed, sometimes their discussion of the classic arguments against ethical naturalism has the ring of an added bonus, as if a simple appeal to Moral Twin Earth is all that is needed to reject ethical naturalism. See, e.g., H&T (1992b), p. 166.

20 This is especially clear with the argument from queerness, since H&T are far more explicit in drawing out its connection with Moral Twin Earth. But we take it that Moral Twin Earth is playing much the same role with H&T's open question argument as well.

that they aren't, their whole project collapses. So, even if our previous criticisms didn't work, H&T still wouldn't have a case against ethical naturalism. The burden of this section is to show that things really are this bad for H&T.

We'll start off, in section 4.1, by pointing out a number of highly misleading features in H&T's thought experiment. Once these are exposed, it becomes clear that Moral Twin Earth doesn't offer any straightforward conclusions about the status of ethical terms. In section 4.2 we'll redescribe the thought experiment in a way that is far less misleading and see what follows. To the extent that this can be done, the "semantic intuitions" that the thought experiment yields are completely on a par with the intuitions that are associated with Putnam's original thought experiment. We conclude that H&T don't even come close to showing that moral terms aren't rigid designators.

4.1 A Flawed Intuition Pump

Thought experiments are a staple of philosophical argumentation, and intuitions are the staple of thought experiments. A good thought experiment, then, is one that elicits firm and reliable intuitions but also intuitions that derive from crucial features of the thought experiment and not from extraneous considerations. The problem is that intuitions are easily manipulated and that the source of the manipulation may be hidden, or tucked away, in what otherwise looks like a straightforward argument. One person who has urged the careful scrutiny of thought experiments–a wonderful storyteller himself–is Daniel Dennett.[21] Dennett calls thought experiments *intuition pumps,* partly to denigrate the argumentative strategy, but largely to bring attention to the sole purpose of a thought experiment, the manipulation of intuitions. He points out that, often enough, the element of a thought experiment that is most responsible for affecting our intuitions is a feature that goes unnoticed, a feature which, once exposed, ought to be dismissed by all hands as irrelevant or misleading.

To take an example, consider a family of thought experiments that are bound to arise in any discussion of whether people have free will. These are scenarios where we are supposed to imagine that we are the puppets of a cosmic puppet master, or prisoners in a cosmic cell, or, to put it generally, where there is some greater, cosmic agent whose control over us deprives us of our own autonomy. Dennett's irreverent discussion brings attention to the way these thought experiments rely upon an agent lurking in the background to make us feel that we cannot be free if determinism is true (Dennett 1984, p. 10).

21 See especially Dennett (1984).

> I cannot prove that none of the bogeymen in this rogues' gallery really exist, any more than I can prove that the Devil, or Santa Claus, doesn't exist. But I am prepared to put on a sober face and assure anyone who needs assuring that there is absolutely no evidence to suggest that any of these horrible agents exists. But of course if any of them did, woe on us! A closet with a ghost in it is a terrible thing, but a closet that is just like a closet with a ghost in it (except for lacking the ghost) is nothing to fear, so we arrive at what may turn out to be a useful rule of thumb: whenever you spy a bogey*man* in a philosophical example, check to see if this scary agent, who is surely fictitious, is really doing all the work.

In other words, one should be highly suspicious of thought experiments that evoke cosmic super agents because the intuitions they support may hinge precisely on the fact that an agent is evoked, even though most parties to the debate will agree that the existence of a such an agent is a nonstarter.

The point of these reflections is that they bear on H&T's appeal to Moral Twin Earth. Like H&T, we will not question the legitimacy of Putnam's original Twin Earth thought experiment. Our question is whether H&T's Moral Twin Earth thought experiment is as legitimate as the original. The whole point of H&T's direct argument is that there is supposed to be an asymmetry between the intuitions generated by Twin Earth and Moral Twin Earth; this asymmetry is supposed to argue for the claim that moral terms aren't rigid designators. For the argument to work, however, the two thought experiments have to be constructed in analogous fashion. The problem with the argument is that they aren't. There are a number of crucial disanalogies between the two thought experiments, and it's these disanalogies that do much of the work in generating the intuitions that H&T's arguments rely upon.

Before we get to the central disanalogies, we note a preliminary problem about the role that H&T assign to causal regulation. Recall that causal regulation is a theory of reference owing to Richard Boyd and that H&T appeal to this particular theory of reference purely as an expository device in describing the difference between Earth and Moral Twin Earth. (On Earth, one functional property causally regulates the term 'good', whereas on Twin Earth, quite a different functional property causally regulates the term.) Notice that H&T build Boyd's theory of reference into the very description of Moral Twin Earth. Yet this isn't at all how the standard Twin Earth thought experiment goes. There one is simply told that XYZ occurs in all of Twin Earth's lakes, streams, and so on. Now H&T claim that the intuition generated by the Moral Twin Earth thought experiment is that twins mean the same thing by "good" and that whatever difference there is between them ought to be attributed to a difference in belief or theory, not a difference in meaning. Their conclusion is that "good" isn't a rigid designator. But given the essential reference to causal regulation, their use

of the thought experiment faces an obvious difficulty. This is the possibility that even if H&T are right about the intuitions, these intuitions may argue more against Boyd's particular theory of reference than against the fundamental claim that "good" and other moral terms are rigid designators.

Still, we don't want to put too much weight on this criticism, in part, because it's very likely that people don't pay attention to H&T's mentioning of causal regulation anyway—a point that should be evident from the discussion in section 3. More importantly, however, we think that other more interesting features of the thought experiment are illicitly at work. We turn to these now.

4.1.1 Competing theories of a kind

In H&T's thought experiment, the way that the contrast is drawn between Earth and Moral Twin Earth is in terms of two competing moral theories—consequentialism and deontology. Both of these theories have their share of plausibility, which is why they both continue to have strong advocates in philosophical circles.[22] On the other hand, in the original Twin Earth thought experiment, XYZ is a philosophical invention. There is no equivocation about this: XYZ is a completely different chemical composition than H_2O, and, moreover, it's a chemical composition that's tied to a chemical theory that no one has ever supposed is true of water. The whole point of talking about XYZ is, as it were, to stipulate that the chemical composition of the stuff that fills their lakes and so on is something with which we have no familiarity.

Recall that H&T's gloss of Moral Twin Earth is that, in contrast with the standard Twin case, our intuitions are that Moral Twin Earthlings aren't referring to different properties with their moral terms; they just have different beliefs and theories about the very same moral properties that our beliefs are about. Yet surely, choosing properties that satisfy a plausible competing theory about moral properties is going to bias the case toward this interpretation. In addition, the situation is especially problematic when the properties in question are poorly understood and when there is little confidence that we have arrived at anything like an adequate account of their nature. This, of course, is how things stand with regard to moral properties, yet emphatically not how they stand with regard to chemical properties like water/H_2O. We have a high degree of understanding when it comes to chemical properties. So here is an important potentially distorting influence in H&T's description of Moral Twin Earth.

22 Notice that much of the disagreement in moral theory just is a disagreement about which of these theories is true.

What would happen if we were to be more careful in describing Moral Twin Earth? We would have to make sure that the Twin case was like the standard water/H_2O case in that the Twin property does not satisfy a competing theory about the nature of the property here on Earth. Maybe the easiest way to avoid this situation is to follow Putnam's lead by merely stipulating the difference. One could say that in all the usual places where one finds such-and-such moral property on Earth, one finds the Z-property on Moral Twin Earth. Then one need only make it clear that the particular theory of Z-ness isn't to be given, except to say that Z-ness is a wholly different functional property than the one found in its place on Earth. We'll describe a case like this in a little while. But for what it's worth, our own intuitions suggest that much of the asymmetry between one's reaction to the original Twin Earth thought experiment and one's reaction to H&T's Moral Twin Earth thought experiment disappears with the single change just mentioned.

4.1.2 Functional and non-functional natural kinds

Another potentially distorting influence on the intuitions about Moral Twin Earth is the fact that moral properties are assumed to be functional properties. In contrast, the original Twin Earth thought experiment is framed in terms of non-functional natural kinds. To gain some insight into whether this asymmetry is relevant, it would help to consider another property or kind that shares this feature with moral properties. Following this strategy, we might consider, for example, how things turn out in a Twin case involving psychological properties. Interestingly, despite the fact that the point of Moral Twin Earth is to pull moral properties apart from other sorts of properties—including the example H&T use as a foil, propositional attitudes—they never actually construct a Twin case involving psychological properties. Rather, they rely completely on the water/H_2O case.[23]

What might Psychological Twin Earth look like? As before, the geography and natural surroundings can be assumed to be almost exactly the same as on Earth. There are analogs there of all the countries and cities on Earth, and analogs of all the humans here as well, and, in general, Psychological Twin Earth is as similar to Earth as it can be, given the following difference. On Earth, judgments and discourse about propositional attitudes are causally regulated by some unique family of functional properties whose essence is (at least in part) functionally characterizable by, say, the

23 It's important to bear in mind that, while discussions of Twin Earth cases inevitably involve questions of what our thoughts and discourse are about, one doesn't actually have a Twin case for psychological properties until they become the *referents* of the terms in question.

generalizations of some automated version of Bayesian decision theory. In contrast, on Psychological Twin Earth, the properties that causally regulate their propositional attitude terms are functional properties whose essence is functionally characterizable by the generalizations of a different decision theory.[24] Again, it doesn't matter if our propositional attitude terms do in fact refer to properties that can be characterized by Bayesian decision theory. The point is just that, whatever sort of functional properties causally regulate propositional attitude terms on Earth, quite different properties do the same job on Psychological Twin Earth. The question, then, is whether propositional attitude terms have different referents in English and Twin English.

Our initial reaction to this case is basically the one that H&T expect. It does seem that the two sets of terms have different referents. Still, it's worth noting that the intuitions seem less secure than the ones generated by the standard Twin case involving H_2O and XYZ. Moreover, the Psychological Twin Earth thought experiment is subject to a number of responses that pose serious challenges to H&T's interpretation of Moral Twin Earth. For instance, one worry about the Psychological Twin Earth story is that there is always the danger that the two psychological theories don't actually provide distinct sets of functional properties. Perhaps there is a common functional core to the two theories. Their differences, then, needn't reflect relevant differences in the functional roles of the properties they pick out. Though the properties would remain unlike one another in certain respects, they could still have a common essential nature.[25] Reflecting on this possibility certainly tends to blur one's intuitions about the case.

Notice as well, if there is a common functional core, then it may follow that Psychological Twin Earthlings have different beliefs and a different theory about the very same types of propositional attitude states that we have beliefs and theories about—the corresponding conclusion that H&T would have us draw for Moral Twin Earth. But in this case, the fact that Earthlings and their twins have different beliefs is of a piece with the common situation where people who inhabit the same environment have different beliefs with respect to a kind. In neither situation does it follow that the terms in question aren't rigid designators. So even if people's intuitions about Moral Twin Earth turned out to be as H&T claim, that wouldn't suffice to show that moral terms aren't rigid designators.

24 We are framing the thought experiment using Boyd's notion of causal regulation, despite the warnings previously mentioned. We'll wait until section 4.2 to make a complete break from H&T's way of framing a twin situation.

25 In a brief yet interesting commentary on H&T (1990/91), Erik Kraemer suggests a response along these lines. See p. 469.

A related worry is that there are no clear criteria for determining how different two functional roles can be while continuing to constitute what is essentially the same functional property. If a functional role for a given propositional attitude were augmented to respect a single new generalization, one that accounts for the state's effects in recherché situations, would that mean that a new propositional attitude has been identified?[26] The moral properties that H&T take as their target are subject to the same sort of difficulty. Yet that's because, by hypothesis, they are functional properties; it's not because they are moral properties. This brings us back to the main point that H&T may gain some false leverage against ethical naturalism merely because, at the crucial point in their argument, they compare ethical properties to non-functional natural kinds like water. Once again, to see whether there is a special problem with moral properties, Moral Twin Earth will have to be redescribed.

4.1.3 The difficulty of isolating moral properties

In evaluating the significance of Moral Twin Earth, one thing to keep in mind is that it's easy to construct a case that yields the intuition that Twin English moral terms have different referents than English terms. All you need to do is make sure that the properties that causally regulate the terms are sufficiently different. For instance, if twin-"good" was causally regulated by a property like temperature (say, the hotter the better), then there would be little problem in supposing that Twin English moral terms have different referents than their English counterparts.

The same point holds for properties that are closer to paradigmatic moral properties. For example, if the term twin-"good" were causally regulated by honesty (instead of goodness), then under the right circumstances one might be compelled to think that twin-"good" has a different meaning than our term despite the superficial similarity. Suppose that we pointed to obvious cases of people who, while honest, were in other respects grossly morally depraved, and suppose that it was perfectly clear that the Twin Earthlings with whom we were discussing the matter recognized the gross moral failings of these people but nevertheless were unfazed and patiently but emphatically pointed out that those features were simply irrelevant to "goodness"; all that matters, they might say, is that the people didn't lie. Under these circumstances, we might decide that these Twin Earthlings were a remarkably strange bunch and that, despite being miraculously like us in every other way, they

26 The point doesn't concern how different an agent's beliefs about a kind can be while remaining constant in its reference—an issue that arises for functional (or conceptual) role theories of meaning. Rather, it concerns how different the functional roles of the kinds themselves can be before they become essentially different.

use one set of terms differently from us, with completely different referents. We say that this might be the conclusion to draw, but the intuitions are admittedly a bit hazy.

Still, the case is useful, as it suggests some other features of H&T's thought experiment that are illicitly at work. One of these stems from the assumption that Moral Twin Earthlings are like their Earthling counterparts in almost every respect. On this assumption, it's extremely natural to suppose that they have some way of referring to all the same sorts of things that we find significant, including moral properties. But if they have the ability to refer to these properties–properties that Earthlings take quite an interest in–there would have to be some special compelling reason to suppose that they did not in fact refer to them. The simple fact that their conceptual system is so much like our own and that they share our broad cultural and social interests–like us, they have governments, rock musicians and so on–is sure to bias the interpretation that their moral terms must refer to the same properties as our own.

A related and even more important consideration is easy to overlook. It's extremely natural to suppose that, if Twin Earthlings are so much like us, that they are *people*. Indeed, it's very hard to describe the scenario without prejudicing the issue. The problem, of course, is that moral theory is supposed to be applicable to (at least) all people. But if moral properties are applicable to beings on Twin Earth, then it would seem that moral properties are instantiated on Moral Twin Earth. This yields another serious disanalogy with the original thought experiment involving water/H_2O. In the standard Twin case, XYZ is said to take the place of H_2O on Twin Earth: wherever H_2O occurs on Earth, XYZ occupies the corresponding place on Twin Earth.[27] But if moral properties occur on Moral Twin Earth (and presumably play much the same roles that they play here), we should expect that the Moral Twin Earthlings have terms for them. The problem is that these sorts of considerations are likely to eclipse the facts in the Twin story about what properties "causally regulate" their use of terms like "good", "wrong", and so on. The business about what causally regulates what is bound to be ignored, given the overwhelming likelihood that beings so similar to us would take an interest in moral properties. Every other property they have lexicalized corresponds exactly to one we have lexicalized. Why stop short of moral properties?

Moreover, it is not even clear that Moral Twin Earth can be coherently described. Many of the features that make Moral Twin Earth familiar and

27 Remember that the standard Twin case is described like this: "Imagine a hypothetical place, Twin Earth, where there isn't any H_2O. Wherever we have H_2O, they have XYZ instead ..."

intelligible–things like the fact that there are countries, and governments, and rock musicians–seem to imply that there are people there. But, again, if there are people there, it's hard to see how H&T are going to be able to enforce the stipulation that Moral Twin Earth is free of the properties that causally regulate Earth's moral terms. The upshot of all of these considerations is that, if our intuitions are to be trusted, Moral Twin Earth has to be specified with far more care than H&T suppose. Unfortunately, it's no simple task to say exactly how the two planets have to differ in order to ensure that the right properties on Moral Twin Earth take the place of Earth's moral properties. At the very least, significant differences in human nature and/or circumstances will certainly be required.

4.2 Moral Twin Earth Revisited

We have highlighted some clear disanalogies between H&T's Moral Twin Earth and Putnam's original thought experiment and have noted that these may illicitly push one's intuitions in a direction that favors H&T's case against ethical naturalism. The question is, once these disanalogies are corrected, does the Moral Twin Earth thought experiment continue to support the conclusion that moral terms aren't rigid designators? We think it does not. To show why, we will now sketch a revised description of Moral Twin Earth, one in which we'll do our best to guard against these disanalogies and other misleading influences.[28]

Let's begin by supposing that on Earth our general knowledge about the world increases to the point where moral properties are well understood and enter into powerful, empirical explanations of various phenomena. At this point, people come to have the same regard for moral theory and for moral properties as they do for chemistry and for chemical properties. Suppose also that one of the discoveries of moral science is that, on Earth, moral properties precisely correspond to a particular class of functional properties; "good", for example, corresponds to the functional property N. When we later discover Moral Twin Earth, it superficially resembles Earth. It looks much the same in that it has rivers where Earth has rivers, and mountains where Earth has mountains, and it even has inhabitants that resemble the inhabitants of Earth in various ways. In particular, there is a group of Twins that seems to speak English; that is, their words sound just like English words and seem to refer to the same sorts of things that our words refer to.

28 It is worth noting again, however, that it may not be possible to give a coherent description of Moral Twin Earth, particularly in light of the fact that there should be no people there. Is it consistent, for example, to suppose that the Twin Earthlings can speak a language basically like English without being people?

However, there are a couple of important differences. These inhabitants, despite resembling us a great deal and despite appearing to speak a language much like English, are not people, and nothing on Moral Twin Earth plays the functional role associated with moral properties on Earth. For instance, nothing on Moral Twin Earth has the functional property N. Moral Twin Earthlings do appear to talk of things as being "good"—they use this form of words—but in so doing their use of the term is guided by a rather different functional property R. To someone with only a superficial understanding of the nature of moral properties on Earth, things that have the functional property R seem much the same as things that have the functional property N. Yet this is not at all true of those people, the moral scientists, who have a facility with moral theory and a clear and full understanding of the nature of moral properties on Earth. In short, the inhabitants of Moral Twin Earth are as much like human beings as is compatible with these differences. In our present state of knowledge, we are unable to say exactly how similar this is.

The crucial question is whether on Moral Twin Earth "good" refers to the functional property R. This is not an easy question—primarily because it is not clear that one really *can* coherently imagine the Moral Twin Earth thought experiment. But to the extent that the thought experiment is coherent, we submit that the natural inference is quite the opposite of the one that H&T have drawn. To our ears, anyway, the natural thing to say is that "good" *does* refer to the functional property R. After all, this is the property that guides the use of their term "good", and this property obviously plays an extremely important role in the lives of the inhabitants of Moral Twin Earth, just as the functional property N plays an extremely important role for us here on Earth. Moreover, the functional property N, which our own term "good" refers to, is not even instantiated on Moral Twin Earth and it plays no role at all in the lives of the inhabitants there. Finally, it is clear to experts—both on Earth and on Twin Earth—that the functional properties N and R are fundamentally different, though they are superficially alike. Under these circumstances, why in the world would we take the inhabitants of Moral Twin Earth to be referring to the functional property N, a property they've never even encountered, rather than the functional property R, which is everywhere around them, which they have a powerful, highly explanatory theory of, and which plays such a pivotal role in their lives? Surely, if the thought experiment is coherent at all, then the "moral" terms of Moral Twin Earth have different referents than our moral terms; they involve a difference in *meaning* and not merely a difference in belief or theory.

Of course, as we've noted, the thought experiment might not get off the ground at all. Taking just the most obvious problem, can we really coherently

imagine beings so much like us who are not people? But if we illicitly take the inhabitants of Moral Twin Earth to be people, then clearly our own moral properties apply to them. When twin-Hitler presides over their Holocaust, his actions are morally wrong in exactly the same way that Hitler's actions were. Our self-same moral properties apply to twin-Hitler's actions as much as they apply to Hitler's. Given the overwhelming importance we attach to these properties, and the equally important role that they will play in the lives of the inhabitants of Moral Twin Earth, it is only natural that we should take the "moral" terms there to denote the same properties as they do for us. And we would assume that any differences of belief between us and our counterparts on Moral Twin Earth about "the good" are just that—differences of *belief*, not of reference. But this is to abandon the project of constructing a thought experiment that is parallel to Putnam's original one. And it's clear that the disanalogies are doing all the work. If Putnam's Twin Earth had both H_2O and XYZ on it, and the H_2O played all the same roles as it does here on Earth, no one would want to say that the twin English word "water" refers to XYZ and not H_2O.[29]

Dennett's caution about thought experiments is more than justified in the present case. Once Moral Twin Earth is described with sufficient care, it fails to support the strong philosophical conclusion with which it was originally associated, and the direct argument no longer works. Moreover, since H&T's attempts to revive the argument from queerness and the open question argument turn on the claim that ethical terms aren't rigid designators, they depend on the success of the direct argument. Since the latter doesn't work, we have yet another reason for thinking that their other arguments don't work either. We conclude that Moral Twin Earth does absolutely nothing to undermine ethical naturalism.

5 Conclusion

H&T's convictions about Moral Twin Earth are firm and clear. They don't just think that Moral Twin Earth offers an interesting philosophical example to mull over. Rather, their claim is that Moral Twin Earth is the basis for a nearly decisive refutation of ethical naturalism. "The new wavers are defending (to borrow terminology from the chess world) a lost position" (H&T 1992b, p. 171).

29 Similar points apply to other disanalogies between H&T's thought experiment and Putnam's. The original asymmetry in our reactions to the two thought experiments may well simply be the result of the fact that moral kinds are, e.g., poorly understood. But then the upshot of the thought experiment is just that moral kinds are poorly understood—hardly a devastating new objection to ethical naturalism.

We think this is all wrong. Not one of the arguments that H&T link to Moral Twin Earth stands up to scrutiny. On the contrary, we've argued, first, that H&T's formulation of Moral Twin Earth fails to reestablish either Mackie's argument from queerness or Moore's open question argument. Second, the Moral Twin Earth thought experiment contains a number of serious disanalogies with Putnam's Twin Earth case, and it's these asymmetries that perform much of the work in generating the intuitions that H&T cite. Third, when the various asymmetries are removed–to the extent that they can be removed–a neutral reader's intuitions aren't the ones that H&T need. Finally, since their whole series of arguments rest upon these intuitions, their conclusion, that ethical naturalism is in trouble, has no support.

A realistic naturalistic view of ethical properties remains a reasonable goal in metaethics. On the one hand, the tendency outside of ethics toward a naturalistic outlook is gaining momentum, with a fair amount of progress in adjacent areas of philosophy, such as semantics and the study of mental representation. On the other hand, as H&T themselves admit, the standard reasons for thinking that ethical properties are inherently different from natural properties have little to be said for them. In the end, the status of ethical naturalism rests on the positive case that can be made in favor of ethical properties. But such an account will surely require a clean slate as its starting point. It's in this spirit that we view our critique of H&T's discussion of Moral Twin Earth. If we are right, then yet another argument against ethical naturalism can be dismissed. Far from being in a "lost position", ethical naturalists may find themselves to be just a few moves into the game.[30]

Stephen Laurence
University of Hull
Philosophy Department
Hull HU6 7RX
s.laurence@phil.hull.ac.uk

Eric Margolis
Rice University
Department of
Philosophy-MS 14
Houston, Texas 77251-1892
margolis@ruf.rice.edu

Angus Dawson
University of Keele
Department of Philosophy
Keele, Staffordshire, ST5 5BG
a.j.dawson@keele.ac.uk

References

Blackburn, S. (1985). "Supervenience Revisited" in I. Hacking, ed., (1985) *Exercises in Analysis: Essays by Students of Casimir Levy,* Cambridge: Cambridge University Press.

Brink, D. (1984). "Moral Realism and Skeptical Arguments from Disagreement and Queerness", *Australasian Journal of Philosophy* 62: 111–25.

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

30 We would like to thank David Phillips for helpful comments on this paper.

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

Dennett, D. (1984). *Elbow Room: The Varieties of Free Will Worth Wanting,* Cambridge, MA: MIT Press.

Frege, G. (1892). "On Sense and Reference" in P. Geach and M. Black, eds., (1952) *Translations from the Philosophical Writings of Gottlob Frege,* Oxford: Blackwell.

Hare, R. M. (1952). *The Language of Morals,* Oxford: Oxford University Press.

Hare, R. M. (1995). "A New Kind of Ethical Naturalism" in P. French, T. Uehling, and H. Wettstein, eds., (1995) *Midwest Studies in Philosophy, XX, Moral Concepts,* Notre Dame, Indiana: University of Notre Dame Press.

Horgan, T., & Timmons, M. (1990/91). "New Wave Moral Realism Meets Moral Twin Earth", *Journal of Philosophical Research* Vol. XVI: 447–65.

Horgan, T., & Timmons, M. (1992a). "Troubles on Moral Twin Earth: Moral Queerness Revived", *Synthese* 92: 221–60.

Horgan, T., & Timmons, M. (1992b). "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived", *Philosophical Papers* Vol. XXI, No. 3: 153–75.

Kim, J. (1984). "Concepts of Supervenience" in his (1993) *Supervenience and Mind,* Cambridge: Cambridge University Press.

Kraemer, E. (1990/91) "On the Moral Twin Earth Challenge to New-Wave Moral Realism", *Journal of Philosophical Research,* Vol. XVI: 467–72.

Kripke, S. (1972). *Naming and Necessity,* Oxford: Blackwell.

Lewis, D. (1986). *On the Plurality of Worlds,* Oxford: Blackwell.

Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong,* Harmondsworth: Penguin.

McLaughlin, B. (1995). "The Varieties of Supervenience" in E. Savellos and Ü. Yalcin, eds., (1995) *Supervenience: New Essays,* Cambridge: Cambridge University Press.

Moore, G. E. (1903). *Principia Ethica,* Cambridge: Cambridge University Press.

Putnam, H. (1970). "On Properties" in his *Philosophical Papers,* vol 1. Cambridge: Cambridge University Press.

Putnam, H. (1973). "Meaning and Reference", *Journal of Philosophy* 70: 699–711.

Putnam, H. (1975). "The Meaning of 'Meaning'" in his *Philosophical Papers,* vol. 2. Cambridge: Cambridge University Press.

Railton, P. (1986). "Moral Realism", *Philosophical Review* 95: 163–207.