G. E. Moore held that all forms of ethical naturalism rested on a fallacy— the so-called ‘naturalistic fallacy’— which his open question argument was designed to expose. For a while anyway, the argument’s persuasive appeal held a good many philosophers in its sway. Its appeal was felt by W. D. Ross (1930: 7-11, 92-3) and A. C. Ewing (1948: 41-2) who, like Moore, were led to espouse ethical nonnaturalism. Its appeal was also felt by A. J. Ayer (1952: 104-05) and R. M. Hare (1952: ch. 5) who, because they found the metaphysical and epistemological commitments of nonnaturalism unpalatable, were led to noncognitivist accounts of moral discourse.1 But as it turned out, the persuasive appeal of Moore’s argument lasted only as long as did the appeal of certain semantical views upon which ethical naturalism was presumed to rest. Once these semantical views were questioned and the philosophical soil was fertilized with more plausible views, a novel strain of ethical naturalism—a kind allegedly immune to Moorean open question arguments—was bound to sprout forth. One now sees this new strain of ethical naturalism everywhere, though we think it first emerged in the hearty intellectual climate of upper New York state.

Our central aim in this paper is to cast doubt on the sort of new wave moral semantics that has recently been pressed into service on behalf of ethical naturalism. We argue that the currently popular version of naturalism, despite its immunity to Moore’s version of the open question argument, succumbs to a newly fashioned open question argument. Since new wave moral semantics is at the heart of the recent strain of
ethical naturalism, we take ourselves to be going for the jugular. We do all this in the second and third sections of the paper. But first, in order to bring things more clearly into focus, let us pause to remember Moore’s open question argument.

I Moore’s Open Question Argument Remembered

Here we take ethical naturalism to be at bottom a metaphysical theory about the existence and nature of putative moral properties (like goodness and rightness) and moral facts involving such properties. The view can be put this way:

**MN** Metaphysical ethical naturalism: There are moral properties (and facts); and these are identical with natural properties (and facts).¹²

One line of support for MN asserts that certain moral terms like ‘good’ have analytically true naturalistic definitions; i.e., these terms are synonymous with natural terms or expressions referring exclusively to processes, states, or properties that are part of the subject matter of the sciences. Call this view ‘analytic semantic (ethical) naturalism’, which can be expressed thus:

**ASN** Analytic semantic naturalism: Fundamental moral terms like ‘good’ have analytically true naturalistic definitions.

Traditionally, MN and ASN have been closely associated; so closely in fact, that one often finds authors assuming that MN and ASN are a package deal.³ Call this package analytic ethical naturalism. The source of this close association seems to have been acceptance of a synonymy criterion of property identity, according to which two property referring expressions refer to the same property if and only if those expressions are synonymous – i.e., have the same meaning. Given this criterion, the truth of ASN is not merely sufficient for the truth of MN, but necessary as well.

Moore’s open question argument was directed at analytic ethical naturalism. Given the assumption that MN requires ASN, Moore supposed that refuting the semantic thesis was sufficient for refuting the metaphysical thesis. His attempted refutation of ASN turned on what he
took to be a crucial test of the theory, a test involving consideration of whether or not a certain form of question was open. A question is open if and only if it is possible for someone to completely understand the question, yet not know its answer; otherwise it is closed. The rationale behind this test is clear. If ASN is true, then statements of the form ‘Anything which has natural property N is good’ are analytically true. And any competent speaker (i.e., anyone who understands the statement, including of course, the meanings of the terms it contains) will know, merely upon attentive, clear-headed, reflection on the terms contained in the statement, that it is true. In other words, for a fully competent speaker, the statement will be knowable a priori. But if statements of the above form are a priori, then surely questions of the form

Q1 Entity e has natural property N, but is it good?

will be closed; any competent speaker will know, upon contemplating the meaning of the question, that if entity e does indeed have N, then the question’s answer is affirmative. Of course, if Q1 is open, the hypothesis about the meaning of ‘good’ expressed in ASN is false.

So Moore defended what we’ll call the ‘open question thesis’:

OQT Open question thesis: questions of the form Q1, for ‘good’ and for other moral terms, are open questions.

Moore claimed, with a good deal of plausibility, that if Q1 were closed, then it should strike us intuitively as being on a par with the question, ‘Entity e has natural property N, but is it N?’ But while this question is utterly trivial and its answer obvious, the same isn’t true of questions having the form of Q1. As Moore remarked in connection with the suggestion that ‘good’ just means ‘pleasure’: ‘whoever will attentively consider with himself what is actually before his mind when he asks the question ‘Is pleasure (or whatever it may be) after all good?’ can easily satisfy himself that he is not merely wondering whether pleasure is pleasant.’ (1903: 16)

It is worth noting that although Moore focused on questions having the form of Q1, there is another kind of question equally relevant to the testing of ASN, viz., the converse of Q1:

Q2 Does entity e, which is good, have natural property N?
If ASN is correct then Q2 should also be closed. But it too is open; i.e., reflecting on meanings of our terms does not lead one to judge that the answer is trivially, 'yes'.

Moore's central argument against analytic ethical naturalism, then, was that MN requires the truth of ASN, but since this semantic thesis fails the open question test, it is false. Thus, Moore concluded, MN is likewise false. Schematically, his argument can be reconstructed this way:

1. MN → ASN
2. ASN → ~ OQT
3. OQT
Therefore,
4. ~ MN

Analytic ethical naturalism does seem to be a sitting duck for Moore's open question argument, but it is fairly clear that the argument fails as a refutation of all forms of ethical naturalism. The most obvious problem with the argument is the first premise. The claim that a moral property is identical with some natural property only if the terms designating these properties are synonymous derives, as we have already mentioned, from the synonymy criterion of property identity. But this criterion (which entails the claim that property identity requires synonymy of expressions designating the properties in question) is pretty implausible in light of numerous counterexamples from the sciences. The (sortal) property being water is identified with the property being composed of H₂O molecules; being a cloud is identified with being a mass of water droplets, temperature is identified with mean kinetic energy, and so on. But no one supposes that 'being water' is synonymous with 'being composed of H₂O molecules', or that 'temperature' is synonymous with 'mean kinetic energy', and so forth for many other scientific identities. Quite simply, it doesn't seem that, in general, property identity requires synonymy.

So it looks as if the best bet for the ethical naturalist is to grant that ASN is implausible but reject the first premise of the above (reconstructed) Moorean argument. The claim, then, is that moral properties (and the facts that involve them) are identical with natural properties even though expressions designating the relevant moral and natural properties involved in the identity are not synonymous. But then
TROUBLES FOR NEW WAVE MORAL SEMANTICS

it is fair to ask just how the ethical naturalist proposes to understand the semantics of moral terms. Enter new wave moral semantics.

II New Wave Moral Semantics

What we are here calling ‘new wave moral semantics’ is a view that has resulted from the attempt to extend relatively recent developments in the philosophy of language to the understanding of moral language. Three developments particularly deserve mention, since they have figured centrally in new wave moral semantics.

First, as mentioned already, in light of all sorts of counterexamples, particularly from the sciences, there has been widespread rejection of a synonymy criterion of property identity. Second, ever since the pioneering work of Kripke (1972) and Putnam (1973), there has been articulation and widespread acceptance of the idea that names and natural kind terms are rigid designators – rigid in the sense that such expressions designate the same entity with respect to every possible world in which that entity exists.5

Two facts are especially important about identity statements involving rigid designators flanking the identity sign. First, typically such statements are necessarily true without being analytic. Second, some such statements, such as ‘Water = H2O’, constitute definitions – not the kind that express meaning connections and thus are analytic, but rather synthetic definitions that give the real nature or essence of the entity, property, or kind designated by a particular term. Thus, if true, ‘Water = H2O’ expresses the real, underlying essence of water and provides us with a (synthetically true) definition of ‘water’.

A third recent trend in philosophy of language is the widespread acceptance of so-called ‘causal’ theories of reference for names and natural kind terms. In the simplest versions, such theories assert that the semantical property of reference is to be understood as essentially involving appropriate causal connections between speakers’ use of a term and the thing to which the term refers. Such theories propose to explain (i) how the reference of a term is originally determined (e.g., there being some sort of baptism or dubbing ceremony through which speakers in causal contact with an item acquire the ability to refer to that item through the use of some expression used in the ceremony), and (ii) how the capacity to refer is spread throughout a linguistic community (again, by speakers’ causally interacting with one another and with the
item). Of course, this rather simple sketch can be elaborated in a number of ways, but the basic idea is clear: for some terms at least, reference is ‘grounded’ by relevant causal hookups between speakers and the world.

New wave moral semantics has emerged in the context of these developments. One finds a representative version of this semantical view in the work of Richard Boyd (1988), who puts the view to work in his recent defense of (a naturalistic version of) moral realism. All three of the above mentioned ideas are present in Boyd’s semantic views about such moral terms as ‘right’ and ‘good’.

First, he contends that ‘good’, like many other terms, has a synthetic or what he calls a ‘natural’ definition that reveals the essence of the property that term expresses. And, of course, this claim implies the rejection of the synonymy criterion of property identity: the property goodness is identical with such and such natural property even though the term ‘good’ is not synonymous with any naturalistic term or phrase designating the relevant natural property.

Second, in claiming that moral terms admit of synthetic definitions, Boyd evidently takes these terms to be rigid. Like natural kind terms, moral terms allegedly rigidly designate the properties (natural properties for the ethical naturalist) to which they refer. As explained above, it is rigidity that underlies the necessity possessed by synthetic definitions.

Third, Boyd maintains that for moral terms, just as for names and natural kind terms, reference is a matter of there being certain causal connections between the use of moral terms and the relevant natural properties. According to Boyd’s own version of the causal theory of reference, reference is essentially an epistemic notion and so the relevant causal connections constituting reference are just those causal connections involved in knowledge gathering activities:

Roughly, and for nondegenerate cases, a term $t$ refers to a kind (property, relation, etc.) $k$ just in case there exist causal mechanisms whose tendency is to bring it about, over time, that what is predicated of the term $t$ will be approximately true of $k$ (excuse the blurring of the use-mention distinction). Such mechanisms will typically include the existence of procedures which are approximately accurate for recognizing members or instances of $k$ (at least for easy cases) and which relevantly govern the use of $t$, the social transmission of certain relevantly approximately true beliefs regarding $k$, formulated as claims about $t$ (again excuse the slight to the use-mention distinction), a
pattern of deference to experts on $k$ with respect to the use of $t$, etc . . . When relations of this sort obtain, we may think of the properties of $k$ as regulating the use of $t$ (via such causal relations) . . . (1988: 195).

Extending this version of the causal theory of reference to moral terms, as Boyd proposes to do, commits him to what we'll call the 'causal regulation thesis':

**CRT**  *Causal regulation thesis:* For each moral term $t$ (e.g., 'good'), there is a natural property $N$ such that $N$ alone, and no other property, causally regulates the use of $t$ by humans.

(We shall say, of such an $N$, that it 'uniquely' causally regulates the use of $t$ by humans.) On Boyd's view then, the fact that humankind's uses of moral terms are regulated in the way described in CRT is what allows one to conclude that moral terms like 'good' behave semantically like natural kind terms: they rigidly refer to certain natural properties and hence possess synthetic definitions. So we can encapsulate what we are calling new wave moral semantics (at least as developed by Boyd) as the following thesis:

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

If CSN is true, then each moral term $t$ should have a synthetically true natural definition whose definiens characterizes, in purely naturalistic vocabulary, the natural property that uniquely causally regulates the use of $t$ by humans.¹⁰ (In addition, CRT is a corollary of CSN, since CSN cannot be true unless each moral term $t$ is indeed causally regulated by some unique natural property $N$.)

Let us call the combination of MN with CSN, *causal ethical naturalism*. This new strain of ethical naturalism entirely avoids Moore's open question argument, since any force that that argument has is only good against ASN, and ASN is no part of new wave ethical naturalism. Nevertheless, Boyd's semantical views seem to commit him to the view that if MN is true then CSN is also true.

Our project here is to develop a new open question argument which, we claim, casts serious doubt on CSN and thereby on MN. For purposes
of the present discussion, we are willing to grant CRT for argument's sake. Our central claim will be this: even if CRT is true, nevertheless moral terms do not rigidly refer to the natural properties that causally regulate their use by humans. Although causal regulation may well coincide with — or even constitute — reference for certain terms (e.g., names and physical natural-kind terms), we claim that for moral terms anyway, causal regulation does not coincide with reference. I.e., moral terms do not refer to the natural properties that (we are supposing) causally regulate their use by humans. A newly formulated version of the open question argument, we claim, supports our main contention.

III Twin Earth Scenarios and Closed Questions

It will be instructive to approach this matter by first considering the analogous issue concerning terms like 'water'. Suppose someone grants that the use of 'water' by humans is causally regulated by some specific physico-chemical natural kind, but then questions the claim that 'water' rigidly designates the natural kind (viz., H₂O) which happens to fill this role. (This skeptic might believe, for instance, that 'water' designates a more general physical natural kind — a genus which has H₂O as only one of its various actual or physically possible species.) What kind of evidence can be put forth to support the contention that 'water' really does rigidly designate the sortal kind-property H₂O?

When philosophers defend such semantic theses with respect to, e.g., names and physical natural kind terms, a particular type of thought experiment looms very large: the Putnam-style Twin Earth scenario. Recall how those go. In one of Putnam's stories, we are to imagine Twin Earth — a planet pretty much like Earth except that the oceans, lakes, and streams are filled with a liquid whose outward, easily observable, properties are just like those of water, but whose underlying physico-chemical nature is not H₂O, but some other molecular structure XYZ. Despite outward similarities and the fact that speakers of twin English apply the word 'water' to this liquid composed of XYZ, reflection on this scenario yields a very strong intuition that Twin Earthlings don't mean by their twin-English term 'water' what we mean by 'water', and that their term is not translatable by our orthographically identical term. And along with this judgment come two further intuitive judgments: (i) that the meaning of the English term 'water' is such that in any possible world w where this term designates at all, it designates the same
physical natural kind in \( w \) that it actually designates on Earth (viz., \( \text{H}_2\text{O} \)); and (ii) that this fact explains why the English and twin-English terms differ in meaning.

Competent speakers have a strong intuitive mastery of both the syntactic and the semantic norms governing their language. Consequently, the intuitive judgments just described concerning the Twin Earth scenario constitute important (though of course defeasible) empirical evidence for the hypothesis that ‘water’ rigidly designates the specific physico-chemical physical kind that happens to causally regulate the use of this term by humans, viz., the kind \( \text{H}_2\text{O} \).\(^{13}\)

The form of argument just canvassed can be called a _semantic competence argument_.\(^{14}\) It is importantly related to Moore-style considerations concerning open and closed questions, as we shall now explain. First, we need a characterization of the closed/open distinction that suitably updates Moore’s. So let 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’).\(^{15}\) The idea is that semantic competence alone, apart from any specific empirical knowledge the speaker might or might not 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 not closed. Here the idea is that semantic competence alone does not yield an intuitive judgment about the question’s answer; those who do form a judgment probably are relying on additional knowledge, as evidenced by their not regarding the answer as obvious in light of how terms are employed.

The statement ‘Water = \( \text{H}_2\text{O} \)’ is held to be a synthetic definition that gives the real essence of water, not an analytic definition that gives a synonym for ‘water’. Since a person linguistically competent with the terms ‘water’ and ‘\( \text{H}_2\text{O} \)’ need not know that the stuff humans call ‘water’ is composed of \( \text{H}_2\text{O} \) molecules, the following two questions are open:

\[
\begin{align*}
\text{Q3} & \quad \text{Liquid } L \text{ is } \text{H}_2\text{O}, \text{ but is it water?} \\
\text{Q4} & \quad \text{Liquid } L \text{ is water, but is it } \text{H}_2\text{O}?
\end{align*}
\]

But the relevant analogs to the Moorean questions Q1 and Q2 are not
Q3 and Q4, but rather the following questions, which have built into them the appropriate empirical hypothesis about causal regulation:

Q5 Given that the use of ‘water’ by humans is causally regulated by the natural kind H₂O, is liquid L, which is H₂O, water?
Q6 Given that the use of ‘water’ by humans is causally regulated by the natural kind H₂O, is liquid L, which is water, H₂O?

Now, a competent speaker who reflects on Q5 and Q6, and who taps into his own linguistic competence by reflecting carefully on Putnam’s Twin Earth scenario, will judge (i) that the twin-English term ‘water’ differs in meaning from the corresponding English term, enough so that the two terms are not intertranslatable; (ii) that the twin human term ‘water’ rigidly designates the physical natural kind that causally regulates its use by twin humans (viz., XYZ); (iii) that the human term ‘water’ rigidly designates the physical natural kind that causally regulates its use by humans (viz., H₂O); and (iv) that in virtue of facts (i)-(iii), the answer to both Q5 and Q6 is obviously ‘yes’. Thus, Q5 and Q6 are closed questions – a fact providing strong empirical evidence in favor of the claim that ‘water’ rigidly designates the specific physical natural kind that causally regulates the use of this term by humans.

IV Troubles for New Wave Moral Semantics: The Open Question Argument Revived

Competent speakers, as we said, have substantial intuitive mastery of the syntactic and semantic norms governing the proper use of terms in their language. Mastery of the semantic workings of ‘water’, for instance, is reflected in people’s strong intuitions about Putnam’s Twin Earth scenario, and in the fact that Q5 and Q6 are closed questions. Presumably, competent speakers have a comparable intuitive mastery of the semantic workings of ‘good’ and other fundamental moral terms. So if causal semantic naturalism is correct, then things should go the same way they go with ‘water’. That is, if indeed the term ‘good’ rigidly designates the unique natural property (if there is one) that causally regulates the use of ‘good’ by humankind in general, then it should be possible to construct a suitable Twin Earth scenario with these features: (i) reflection on this scenario generates intuitive judgments that are comparable to those concerning Putnam’s original scenario; (ii) these judgments are
accompanied by the more general intuitive judgment that 'good' does indeed work semantically as CSN says it does; and (iii) in light of this latter judgment, the relevant analogs of Q5 and Q6 are closed questions.

Conversely, if the appropriate Twin Earth scenario does not have these features — i.e., if the semantic intuitions of competent speakers turn out not to be what they should be if CSN is true — then this will mean that CSN is probably false. We say ‘probably’ false because the inference to CSN’s falsity would be inductive, an inference to the best explanation. Speakers’ semantic intuitions about Twin Earth scenarios are empirical evidence about matters of semantics (just as their syntactic intuitions about grammaticality are empirical evidence about matters of syntax). Such intuitions, especially when robustly present among most all competent speakers, are quite powerful evidence; ceteris paribus, a semantic hypothesis that respects the intuitions is preferable to, and is more likely to be correct than, a semantic hypothesis that repudiates them.16

We plan to mount an argument against CSN (and MN) by arguing that things go the latter way — i.e. one’s intuitive judgments concerning a suitable Twin Earth scenario go contrary to CSN, and the following two questions are both open:

Q7  Given that the use of ‘good’ by humans is causally regulated by natural property N, is entity e, which has N, good?

Q8  Given that the use of ‘good’ by humans is causally regulated by natural property N, does entity e, which is good, have N?

What is wanted is a Twin Earth where things are as similar to Earth as possible, consistent with the hypothesis that twin-moral terms are causally regulated, for twin humans in general, by certain natural properties distinct from those natural properties which (as we are here granting for argument’s sake) regulate the use of moral terms by humans in general.

So let’s begin by supposing that, as Boyd maintains, human uses of ‘good’ and ‘right’ are regulated by certain functional properties; and that, as a matter of empirical fact, these are consequentialist properties whose functional essence is captured by some specific consequentialist normative theory; call this theory T.17 We further suppose that there is some reliable method of moral inquiry which, if properly and thoroughly employed, would lead us to discover this fact about our uses of moral terms.
Now consider Moral Twin Earth, which, as you might expect, is just about like good old Earth: same geography and natural surroundings, people who live in the twin United States by and large speak twin English; there is a state they call ‘Tennessee’ that is situated directly south of a state they call ‘Kentucky’; and every year a fairly large number of Twin Earthlings make a pilgrimage to Twin Memphis to visit the grave site of Twin Elvis. You get the idea. Of particular importance here is the fact that Moral Twin Earblings have a vocabulary that works much like human moral vocabulary; they use the terms ‘good’ and ‘bad’, ‘right’ and ‘wrong’ to evaluate actions, persons, institutions and so forth (at least those who speak twin English use these terms, whereas those who speak some other twin language use terms orthographically identical to the terms for good, etc., in the corresponding Earthian dialects). In fact, were a group of explorers from Earth ever to visit Moral Twin Earth they would be strongly inclined to translate Moral Twin Earth terms, ‘good’, ‘right’ and the rest as identical in meaning to our orthographically identical English terms. After all, the uses of these terms on Moral Twin Earth bear all of the ‘formal’ marks that we take to characterize moral vocabulary and moral practice. In particular, the terms are used to reason about considerations bearing on Moral Twin Earthling well-being; Moral Twin Earthlings are normally disposed to act in certain ways corresponding to judgments about what is ‘good’ and ‘right’; they normally take considerations about what is ‘good’ and ‘right’ to be especially important, even of overriding importance in most cases, in deciding what to do, and so on.

Let us suppose that investigation into twin English moral discourse and associated practice reveals that their uses of twin-moral terms are causally regulated by certain natural properties distinct from those that regulate English moral discourse. 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 $T^d$. These functional properties are similar enough to those characterizable via $T^c$ to account for the fact that twin-moral discourse operates in Twin Earth society and culture in much the manner that moral discourse operates on Earth. (We’ve already noted that if explorers ever visit Moral Twin Earth, they will be inclined, at least initially, to construe Moral Twin Earthlings as having beliefs about good and right, and to
translate twin English uses of these terms into our orthographically identical terms.) The differences in causal regulation, we may suppose, are due at least in part to certain species-wide differences in psychological temperament that distinguish Twin Earthlings from Earthlings. (For instance, perhaps Twin Earthlings tend to experience the sentiment of guilt more readily and more intensively, and tend to experience sympathy less readily and less intensively, than do Earthlings.) In addition, suppose that if Twin Earthlings were to employ in a proper and thorough manner the same reliable method of moral inquiry which (as we are already supposing) would lead Earthlings to discover that Earthling uses of moral terms are causally regulated by functional properties whose essence is captured by the consequentialist normative theory Tc, then this method would lead the Twin Earthlings to discover that their own uses of moral terms are causally regulated by functional properties whose essence is captured by the deontological theory Td.

Given all these assumptions and stipulations about Earth and Moral Twin Earth, what is the appropriate way to describe the differences between moral and twin-moral uses of ‘good’ and ‘right’? Two hermeneutic options are available. On the one hand, we could say that the differences are analogous to those between Earth and Twin Earth in Putnam’s original example, to wit: the moral terms used by Earthlings rigidly designate the natural properties that causally regulate their use on Earth, whereas the twin-moral terms used by Twin Earthlings rigidly designate the distinct natural properties that causally regulate their use on Twin Earth; hence, moral and twin-moral terms differ in meaning, and are not intertranslatable. On the other hand, we could say instead that moral and twin-moral terms do not differ in meaning or reference, and hence that any apparent moral disagreements that might arise between Earthlings and Twin Earthlings would be genuine disagreements — i.e., disagreements in moral belief and in normative moral theory, rather than disagreements in meaning.

We submit that by far the more natural mode of description, when one considers the Moral Twin Earth scenario, is the second. Reflection on the scenario just does not generate hermeneutical pressure to construe Moral Twin Earthling uses of ‘good’ and ‘right’ as not translatable by our orthographically identical terms. But if CSN were true, and the moral terms in question rigidly designated those natural properties that causally regulate their use, then reflection on this scenario ought to generate intuitions analogous to those generated in Putnam’s original
Twin Earth scenario. I.e., it should seem intuitively natural to say that here we have a difference in meaning, and that twin English ‘moral’ terms are not translatable by English moral terms. But when it comes to characterizing the differences between Earthlings and twin Earthlings on this matter, the natural-seeming thing to say is that the differences involve belief and theory, not meaning.\textsuperscript{22}

One’s intuitions work the same way if, instead of considering the Moral Twin Earth scenario from the outside looking in, one considers how things would strike Earthlings and Twin Earthlings who have encountered each other. Suppose that Earthlings visit Twin Earth (or vice versa), and both groups come to realize that different natural properties causally regulate their respective uses of ‘good’, ‘right’, and other moral terms. If CSN were true, then recognition of these differences ought to result in its seeming rather silly, to members of each group, to engage in inter-group debate about goodness – about whether it conforms to normative theory $T^c$ or to $T^d$. (If, in Putnam’s original scenario, the two groups learn that their respective uses of ‘water’ are causally regulated by different physical kind-properties, it would be silly for them to think they have differing views about the real nature of water.) But such inter-group debate would surely strike both groups not as silly but as quite appropriate, because they would regard one another as differing in moral beliefs and moral theory, not in meaning.\textsuperscript{23}

Since semantic norms are tapped by human linguistic competence, and since the relevant linguistic competence is presumably reflected in one’s intuitive judgments concerning Twin Earth scenarios, this outcome constitutes strong empirical evidence against CSN.\textsuperscript{24} And since it is a highly non-trivial issue whether the basic good-making natural property is the one (if there is just one) that causally regulates the use of ‘good’ by humans, or instead is the one that causally regulates this term’s use by twin-humans, or instead is some natural property distinct from either of these, the outcome of the Moral Twin Earth thought experiment also undergirds the following ‘revised open question thesis’:

\begin{itemize}
\item ROQT Revised open question thesis: Questions of the form Q7 and Q8, for ‘good’ and for other moral terms, are open questions.
\end{itemize}

But this thesis ought to be false if CSN is true. So we have arrived at an open question argument against causal semantic naturalism similar to
the one we take Moore to have given against analytic semantic naturalism:
(1) CSN → ¬ ROQT; (2) ROQT; therefore, (3) ¬ CSN.

What is the import of all this for MN, the thesis of metaphysical ethical naturalism? In addressing this question one should notice that CSN is a species of the following more generic semantic thesis:

SSN Synthetic semantic naturalism: Fundamental moral terms like 'good' have synthetic naturalistic definitions.

Now it might be thought that even if CSN is not tenable because of Moral Twin Earth, there remains a serious possibility that some other species of SSN still might be tenable – perhaps a version of SSN that nobody has yet articulated. But Moral Twin Earth is more than a specific thought experiment directed at the specific semantic thesis CSN. It is, in addition, a recipe for thought experiments. For any potential version of SSN that might be proposed, according to which (i) moral terms bear some relation R to certain natural properties that collectively satisfy some specific normative moral theory T, and (ii) moral terms supposedly refer to the natural properties to which they bear this relation R, it should be possible to construct a Moral Twin Earth scenario suitably analogous to the one constructed above – i.e., a scenario in which twin-moral terms bear the same relation R to certain natural properties that collectively satisfy some specific normative theory T', incompatible with T. The above reasoning against CSN should apply, mutatis mutandis, against the envisioned alternative version of SSN.

So Moral Twin Earth provides strong grounds for rejecting not only CSN, but also the more generic thesis SSN. The generic scope of Moral Twin Earth also undergirds the following 'general revised open question thesis':

GROQT General revised open question thesis: Questions analogous in form to Q7 and Q8, for 'good' and for other moral terms, are open questions.

(A question will be analogous in form to Q7 or Q8, in the intended sense, if its 'given'-clause mentions, in place of Boyd’s posited relation of causal regulation, whatever relation R supposedly grounds the reference of moral terms according to the particular version of SSN one
is considering.) But if there is any version of SSN that is true, then the corresponding instance of ROQT should be false. We have thus arrived at the following open question argument against metaphysical ethical naturalism, a version that plugs the hole left open by Moore's original open question argument:

1. MN → (ASN v SSN)
2. ASN → ¬ OQT
3. SSN → ¬ GROQT
4. OQT
5. GROQT
Therefore,
6. ¬ MN

V Conclusion

No doubt various defensive stratagems are possible, in response to this postmodern version of the open question argument. The principal options, listed in order of increasing retreat against the force of Moral Twin Earth, appear to be the following:

1. **Bold denial**: Claiming that the Moral Twin Earth thought experiment does not describe a genuinely possible scenario.

2. **Avoidance**: Claiming that although people's semantic intuitions about Moral Twin Earth are not suitably analogous to their intuitions about Putnam's original Twin Earth scenarios, these semantic intuitions are just mistaken and hence do not really undermine causal semantic naturalism at all. After all, brute intuitions don't settle the matter by themselves; they can be overruled by a semantical theory that best accommodates all relevant data, including data about the workings of moral discourse and inquiry. Once all relevant data are considered, the new waver might argue, Boydian semantics has a lot going for it and recalcitrant intuitions should be dismissed.

3. **Betting on the future**: Granting that Moral Twin Earth and the revised open question thesis refute Boyd-style causal semantic naturalism, but claiming that there may yet be some other
version of synthetic semantic naturalism that does not succumb to a similar Moral Twin Earth scenario.

(4) **Relativism:** Granting the above open question argument against MN, and retreating to a relativist version of naturalism; i.e., claiming that although there are no moral properties or moral facts *simpliciter*, nevertheless (i) there are moral properties (and facts) *relative to a person or social group*; and (ii) that these relativized properties (and facts) are natural properties. (One might claim, for instance, that on Earth the terms ‘good’, ‘right’, etc. refer nonrigidly to functional properties whose essence is captured by the consequentialist normative theory $T^c$, whereas on Moral Twin Earth these terms refer nonrigidly to different functional properties whose essence is captured by the deontological theory $T^d$.)

Option (1) incurs an overwhelmingly heavy burden of proof, because of two interrelated facts. First, in order for a Twin Earth thought experiment to serve its intended purpose, the relevant scenario need not be a genuine *metaphysical* possibility, but only a broadly *conceptual* possibility. (Suppose, for instance, that Putnam’s original Twin Earth scenarios turn out to be metaphysically impossible: i.e., no physically possible substance other than H$_2$O can have all the features that Putnam attributes to the mythical natural kind XYZ. This outcome would not alter the *conceptual* coherence of Putnam’s thought experiments, and hence would not alter their relevance to the semantics of ‘water’.) Second, the *prima facie* intelligibility of the above Moral Twin Earth story constitutes very strong evidence that the scenario is indeed conceptually possible – whether it is metaphysically possible or not.$^{26}$

Option (2) also carries an enormous burden of proof. Admittedly, brute intuitions of the sort revealed by Moral Twin Earth are defeasible; thus people’s semantic intuitions might be just mistaken. But the ethical naturalist who claims that they are has some hard explaining to do. First, we are talking about not just any old intuitively strong beliefs involving morality, but intuitive beliefs about *meaning*, surely an important kind of data when one is considering theoretical accounts of meaning, not easily dismissed.$^{27}$ Second, even allowing that such intuitions could be overridden by an otherwise plausible semantics of moral discourse, nevertheless, *ceteris paribus*, a semantical account that respects the
meaning-intuitions of competent language users is to be preferred over one that does not. Third, and most important for present purposes, the evidential import of such intuitions becomes stronger – they become more than mere isolated, brute data bearing on semantical views – if they can be embedded in a wider semantic theory that provides them with a strong rationale. We think semantic theories that provide just such a rationale can be developed.

Although this is not the place to elaborate such a theory, the sort of rationale any such theory might be expected to provide for our meaning-intuitions can be briefly outlined as follows. (1) One of the defining characteristics of a moral code is that it performs an action-guiding role for members of the community in which it is in force. (2) This normative aspect thus amounts to a semantic constraint for interpreting the practices of a community as moral practices, and so is plausibly taken to be built in to the meaning of moral terms like ‘good’ and ‘right’.28 (3) This action-guiding, normative feature of the meanings of moral terms helps explain why our intuitions engaged by the Moral Twin Earth thought experiment go the way they do: essential to the meaning of moral terms like ‘good’ and ‘right’ is their action-guiding function, which both Earthian usage of moral terms and twin Earthian usage of moral terms share. Hence, despite the fact that the use of moral terms by the two groups is regulated by different natural properties, the (orthographically identical) moral terms of the two groups mean the same, contrary to Boydian semantics and consistent with the results of reflection on questions Q7 and Q8. With such an explanation available for why our intuitions go the way they do, the avoidance option lacks any genuine plausibility.

Option (3) looks unreasonably optimistic, for two reasons. First, no interesting version of synthetic semantic naturalism, significantly different from Boyd-style causal semantic naturalism, is currently even remotely in sight. Second, as explained at the close of section 4, there are already strong grounds for believing that any such version of SSN would indeed succumb to its own Moral Twin Earth scenario, and thus to its own revised open question argument.

Option (4) will be very unattractive to new wave meta-ethical naturalists like Boyd. These philosophers espouse moral realism, an adamantly non-relativist position. And in any case, relativism is the last defensive ditch against Moral Twin Earth. If ethical naturalism gets forced into that ditch, there it deserves to die.
Grim as the situation looks for new wave naturalism, resourceful moves are to be expected – especially from the dialectically vigorous region of upstate New York, from whence the latest strain of moral realism first began to spread like ragweed. But in the end, all defensive strategies are likely to prove futile against Moral Twin Earth. The new wavers are defending (to borrow terminology from the chess world) a lost position.29

NOTES

1. The writers just mentioned were all somewhat less sanguine about the use of the open question argument, and some, most notably Hare, only endorsed a suitably modified version of the argument. Still, for all of the above-mentioned moral philosophers, their rejection of ethical naturalism was due in large part to Moorean open question considerations.

2. Naturalists hold that moral properties (and facts) are a species of natural properties (and facts). Materialistically minded naturalists don't necessarily hold that they are physical properties in a narrow sense, properties of the kind posited by physics. Instead, they can hold that moral properties are (identical to) properties that are realized by, or constituted by, physical properties and whose essence is characterizable in non-normative terminology (though not necessarily directly in the terminology of physics). So even realization/constitution versions of ethical naturalism are presumably committed to an identity thesis, involving identities between moral properties and properties whose essence is characterizable in non-normative terms. Thus, such views are not really rivals to identity views. Rather, given that moral properties are natural properties, there is the further metaphysical issue about whether all natural properties are identical with physical properties (in the narrow sense).

3. It has been fairly standard to take ethical naturalism as a thesis about the meaning of moral terms, with very little care taken to sort out metaphysical ethical naturalism and related semantic views like analytic semantic (ethical) naturalism. See, for example, Brandt (1959: ch. 7).

4. If the question has any empirical assumptions, then knowing the answer amounts to knowing what answer would be correct if the assumptions were all true.

5. One important source of support for the rigidity view is Putnam's famous Twin Earth thought experiments. More on these below.

6. See Boyd (1988: 194-5, 209-12). Boyd contrasts natural definitions – definitions of e.g., natural kinds that purport to express the scientific essence or nature of the kind that is the referent of the term – with conventional definitions – the sort of meaning-giving definitions that purport to be analytically true. The distinction between synthetic or natural definitions on the one hand, and analytic or conventional definitions on the other, corresponds to the traditional distinction (made by Locke, for instance) between 'real' and 'nominal' definitions.

7. Boyd's own suggestion about a correct natural or synthetic definition of 'good' is in terms of what he calls 'homeostatic consequentialist properties' of actions, institutions, and so forth. Of course, even if the suggestion is incorrect, this would not necessarily affect his central claim, viz., that the term 'good' admits of a synthetic definition.

8. Boyd's view, as we understand it, is that the essence of goodness, rightness, and other moral properties consists primarily in certain relational connections they bear to one another and to other properties, actions, institutions, and so forth. So we take it that for Boyd, moral properties are functional properties whose essential relations and interconnections are captured by some specific normative moral theory. It is an empirical question, on this view, which normative theory has this status. (Boyd thinks that the relevant theory is probably some version of consequentialism.) This
functionalist construal of moral properties is articulated more explicitly by Brink (1984, 1989), whose version of moral realism is quite similar to Boyd’s.

9. Admittedly, Boyd is not as explicit as one would like about the rigidity of moral terms. But this view fits well with his claim that moral terms have ‘natural’ definitions, and also helps fend off a form of relativism which, of course, runs counter to the realist form of ethical naturalism Boyd defends. We briefly take up the rigidity issue in note 17, and again in section 5.

10. In the present context, the notion of purely naturalistic vocabulary should be understood broadly enough to encompass causal/functional vocabulary. Boyd, as we understand him, holds that moral terms rigidly designate certain functional properties, whose essence is their causal role in a system of causally interrelated properties. Cf. note 7, and section 4.

11. Of course, even if one grants that ‘good’ refers to a natural property N only if N causally regulates the use of this term, it can be questioned whether any genuine natural property really does play such a role for ‘good’ or other moral terms. Not all natural-kind terms succeed in tracking a natural property — as illustrated by terms like ‘caloric’ and ‘phlogiston’. Also, even if one grants causal regulation of moral terms by natural properties, it is still quite contentious whether any single natural property causally regulates the use of ‘good’ for humankind in general; likewise for other moral terms. But here we are granting this highly optimistic assumption because we want to show that even if it is true, CSN is incorrect anyway.

12. Does this claim entail that causal theories cannot provide a fully general account of reference? Not necessarily. That depends upon what one requires of such an account. The claim does entail that in general, even if an entity e bears to a term t the relation Boyd calls ‘causal regulation,’ this is not sufficient for t to refer to e; hence, the reference relation cannot be simply identical to Boyd’s (putative) causal-regulation relation. However, reference still might be just a species of causal regulation; this could be true if moral terms don’t really refer at all, either to properties that might happen to causally regulate them or to any other properties.

13. Some philosophers who espouse causal theories of reference, and who hold that statements like ‘Water = H₂O’ constitute synthetic definitions, tend to regard the use of thought experiments and appeals to intuition as part of an outmoded, unduly aprioristic, philosophical methodology. This tendency is both ironic and misplaced: ironic, because of the key role that Twin Earth thought experiments have played in convincing the philosophical community that names and natural-kind terms are rigid designators; and misplaced, because thought experiments and speakers’ intuitions about them often constitute an important kind of empirical evidence concerning philosophical theses about language and about language/world relations.

14. This label is intended to suggest a relevant analogy between such arguments and a common form of reasoning within empirical linguistics: viz., the appeal to speakers’ intuitions about the grammaticality and/or syntactic ambiguity of certain sentence-like strings of words, as evidence for or against various empirical hypotheses about natural-language syntax. The latter kind of argument rests on the (empirically plausible) background assumption that syntactic intuitions normally reflect what Noam Chomsky has called speakers’ ‘linguistic competence.’ For further discussion of semantic competence arguments and their relation to the methodology of linguistics, see Horgan and Graham (1990).

15. The proviso mentioned in note 4 applies again here. Also, we are construing a question as an interrogative sentence of a particular language; and we are talking about semantically competent speakers of that language.

16. We will return to this dialectical point in section 5, in connection with the second of four defensive stratagems we will consider for trying to evade the revived ‘open question argument’ we present in the present section.

17. As we explained in note 8, we find it natural to construe Boyd as a moral functionalist, as is Brink. This construal will be built into the specific Twin Earth thought experiment we shall now set forth. But if it should turn out that Boyd is better interpreted some other way — say, as holding that moral terms (nonrigidly) designate, in a given socio-cultural situation, whatever first-order physical properties (or property clusters) happen to collectively satisfy Te (in that situation) — then our Twin
Earth story could be modified appropriately. This would not change the moral we shall draw from the story.


19. Since standard deontological normative theories are internally consistent and conceptually cogent, there is no particular reason to doubt that such a theory characterizes a family of genuine functional properties. (For discussions of how functional properties whose essence is captured by a theory are explicitly characterizable by means of (the Ramsey sentence of) that theory, see Block (1978) and Horgan (1984); these papers adapt a format originally proposed by Lewis (1972).) And since deontological functional properties are reasonably similar, overall, to the kinds of functional properties characterized by consequentialist normative theories, there is also no particular reason to doubt that deontological functional properties are physically realizable by certain first-order physical properties (or by certain ‘homeostatic clusters’ of first-order physical properties).

20. Those who were raised Catholic, as we both were, should have little difficulty envisioning this kind of psychological temperament vis-a-vis matters moral. Indeed, we doubt that there is really any single characteristic temperament — any single profile of sentiments — that operates, for Earthlings generally, in matters of morals. But for present purposes one can suppose there is. This supposition fits naturally with the optimistic (though implausible) empirical assumption, which we are already granting for argument's sake, that there is some single set of natural properties that causally regulate the use of moral terms by Earthlings generally. (One could bring the Moral Twin Earth scenario down to Earth by supposing, more realistically but less congenially with Boyd's assumption, that the kinds of temperamental differences we describe actually obtain among different Earthlings themselves.)

21. It should be stressed that differences in normative moral theory, between Earthlings and Twin Earthlings, do not constitute different claims about which property is identical to goodness, or to rightness, etc. For, normative theories do not make such property-identity claims. Rather, they make claims, for instance, about which natural property is the fundamental good-making property, which is the fundamental right-making natural property, etc. Normative theories per se are neutral between the meta-ethical claims (i) that moral properties are identical with these natural properties; (ii) that moral properties are non-natural properties that supervene upon the natural ones without being identical to them; or (iii) that moral properties do not exist at all.

22. Some time after we thought up the Moral Twin Earth story, we discovered that Hare uses a similar story to criticize ethical naturalism (though, of course, his target is analytic ethical naturalism). He has his readers imagine a group of missionaries landing on a cannibal island and discovering that by sheer coincidence 'good' in Cannibalese is apparently a correct translation of 'good' in English. However, whereas the missionaries apply the English term to people of genteel spirit, the cannibals use their term 'good' to commend people who, among other things, collect more scalps than the average. Hare finds it natural to interpret disagreements between missionaries and cannibals as disagreements in the standards used by these different groups in evaluation rather than as mere disagreements in the meanings of the English and cannibal uses of 'good'. He writes:

Even if the qualities in people which the missionary commended had nothing in common with the qualities which the cannibals commended, yet they would both know what the word 'good' meant. If 'good' were like 'red', this would be impossible; for then the cannibals' word and the English word would not be synonymous. If this were so, then when the missionary said that people who collected a lot of scalps were good (cannibal), they would not be disagreeing, because in English (at any rate missionary English), 'good' would mean among other things 'doing no murder', whereas in the cannibals' language 'good' would mean something quite different, among other things 'productive of maximum scalps' (Hare 1952: 148-9).

23. At any rate, this is how they would regard the matter insofar as they rely on their pretheoretic semantic intuitions. Intuitions can become skewed for those who are sufficiently in the grip of a philosophical theory. Some new-wave moral realists, for instance, may by now have become so
strongly gripped by the Boydian conception of moral reference as causal regulation by natural properties that their own intuitions about Moral Twin Earth actually fall into line with their intuitions about Putnam's original case. For them some philosophical therapy may be necessary, to get them back into touch with their semantically true selves.

24. It is worth noting that the Moral Twin Earth thought experiment does not depend for its outcome on the hypothetical assumption that a consequentialist normative moral theory characterizes the essence of the functional properties that (as we are granting for argument's sake) causally regulate human moral judgments. This was only for concreteness and vividness. It makes no substantive difference if one supposes instead that the functional essence of the causally regulating properties is characterized, say, by some deontological theory, or by some mixed consequentialist/deontological theory. Rather, the crucial features of the thought experiment are (i) that the twin-moral judgments of Twin Earthlings track distinct functional properties than those allegedly causally tracked by the moral judgments of Earthlings; (ii) that twin-moral judgments made by Twin Earthlings coincide largely, but not completely, with moral judgments that would be made by Earthlings; and (iii) that twin-moral and moral judgments, respectively, play comparable roles in the social and institutional lives of Twin Earthlings and of Earthlings, respectively.

25. New wavers are fond of claiming that features of our moral discourse and moral inquiry rest on realist assumptions and so demand some sort of realist metaphysics and associated semantics. See Brink (1989: ch. 2) and Boyd (1988). But see Blackburn (1980, 1984, ch. 6), Copp (1991: 613-14) and Timmons (forthcoming) for doubts about this claim.

26. In addition, we see no particular reason to doubt that Moral Twin Earth is metaphysically possible anyway. Cf. note 17.

27. At the very least, the ethical naturalist who would go the avoidance route must plausibly explain (i) why people's meaning-intuitions about moral terms are so strong and so widespread even though they are allegedly mistaken, and (ii) why those intuitions don't work the same way they do in Putnam's original cases.

28. Exactly how a semantic theory should explain this built-in normative dimension of moral terms is not obvious. Ethical internalism, the semantic thesis according to which (roughly) the meaning of moral terms is such that it is not possible (consistent with what moral terms mean) for anyone to judge that some action is, say, wrong, and yet fail to be motivated (or have reason) to avoid performing that action, has been quite typical of semantic views that attempt to incorporate the action-guiding feature into a semantic story about moral discourse. But as critics like Brink (1989: ch. 3) have pointed out, because this sort of local internalist thesis would require, for each and every member of a community who was competent with moral language, that she be relevantly motivated (or have reason for action) on each occasion of her sincere utterance of a moral judgment, it is unduly strong in dismissing the conceptual possibility of the amoralist. But the intrinsic action-guiding feature of moral terms need not be understood in this rather crude way. For one thing, one might opt for a global form of internalism that would allow for all of the members of a community to be amoralists some of the time, and for some of the members to be amoralists all of the time, but would only deny that all of the members could be amoralists all of the time. Global internalism would tie the action-guiding feature of moral discourse to the activities and motivations of the members of a community in a global, community-wide way. (The global internalist might handle the case of the amoralist by adapting, mutatis mutandis, the treatment by David Lewis (1980) of scenarios involving people who undergo so-called 'mad pain'.)

29. This paper is fully collaborative; order of authorship is alphabetical. An earlier version was presented at the Mark Overvold Memorial Conference at St. Olaf College, October, 1990, where Erik Kraemer commented. It was also presented at the 1991 APA Pacific division meetings in San Francisco (Philip Clarke commented), and at the 1991 Spring Colloquium (by Terence Horgan). We thank our conference commentators and also Tom Carson, Seumas Miller, Mark Sainsbury, Nick Sturgeon, Corliss Swain, and John Tienson for their helpful comments. We dedicate the paper to the memory of Mark Overvold.
REFERENCES


