

Terence Horgan and Mark Timmons  
New Wave Moral Realism Meets Moral Twin Earth  
Journal of Philosophical Research  
Volume 16- 1991  
pp. 447 – 465

## **Terence Horgan and Mark Timmons**

### **New Wave Moral Realism Meets Moral Twin Earth**

MEMPHIS STATE UNIVERSITY

**Abstract:** There have been times in the history of ethical theory, especially in this century, when moral realism was down, but it was never out. The appeal of this doctrine for many moral philosophers is apparently so strong that there are always supporters in its corner who seek to resuscitate the view. The attraction is obvious: moral realism purports to provide a precious philosophical good, viz., objectivity and all that this involves, including right answers to (most) moral questions, and the possibility of knowing those answers. In the last decade, moral realism has re-entered the philosophical ring in powerful-looking naturalistic form. In this paper we provide a dialectical overview: we situate the new wave position itself, and also our objections to it, in the context of the evolving program of philosophical naturalism in 20th century analytic philosophy. We seek to show that although this new contender might initially look like championship material, it succumbs to punches surprisingly similar to those that knocked out the old-fashioned versions of naturalist moral realism.

#### **1. Introduction**

There have been times in the history of ethical theory, especially in this century, when moral realism was down, but it was never out. The appeal of this doctrine for many moral philosophers is apparently so strong that there are always supporters in its corner who seek to resuscitate the view. The attraction is obvious: moral realism purports to provide a precious philosophical good, viz., objectivity and all that this involves, including right answers to (most) moral questions, and the possibility of knowing those answers. In the last decade, moral realism has re-entered the philosophical ring in powerful-looking naturalistic form. Developments in other areas of philosophical inquiry have helped rejuvenate this position, by apparently providing for a strain of naturalism which is immune to the blows that decked older versions

— 448 —

of moral realism, including traditional versions of ethical naturalism. This new wave moral realism has come to dominate recent work in metaethics.

However, despite any advantages this new strain of naturalistic moral realism enjoys over previous versions of realism, we maintain that it is destined to travel the same road as its predecessors. In a series of papers (Horgan and Timmons 1991, forthcoming; Timmons 1990) we have made our case for this pessimistic prediction,

investigating at some length the difficulties faced by new wave moral realism. In this paper we provide a dialectical overview: we situate the new wave position itself, and also our objections to it, in the context of the evolving program of philosophical naturalism in 20th century analytic philosophy. We seek to show that although this new contender might initially look like championship material, it succumbs to punches surprisingly similar to those that knocked out the old-fashioned versions of naturalist moral realism. New-wave realism too has a glass jaw.

## 2. Philosophical Naturalism

Moral realism faces well-known, serious philosophical obstacles- obstacles that seem particularly difficult given a thoroughgoing philosophical naturalism. So we begin the saga of ethical naturalism with some remarks about philosophical naturalism.

We take the naturalist outlook in philosophy to be at bottom a metaphysical view about the nature of what exists. The vague, pre-theoretic idea that this view attempts to articulate and defend is simply the idea that all that exists—including any particulars, events, properties, etc.—is part of the natural physical world that science investigates. This core idea can be precisified in various ways, but common to most variants is what we'll call the *thesis of physical ontic primacy*:

(N1) The ontologically primary or fundamental entities (particulars, properties, facts, etc.) in the world are all part of the subject matter of physics.

There is more to metaphysical naturalism than thesis (N1), however. Common sense, and ordinary discourse, recognize all sorts of entities that seem, at least *prima facie*, not to be literally identical with entities that are *narrowly physical*, i.e., explicitly countenanced by physics. And this fact generates a broad project for the philosophical naturalist: *viz.*, to somehow accommodate, within a broadly naturalistic metaphysical/epistemological framework, *all* particulars, properties, facts, etc. as part of the “natural order.” The operative notion of accommodation is itself vague and pre-theoretic. It too can be precisified in various ways, but the leading idea is what we'll call the thesis of *explainability*:

(N2) All truths are ultimately explainable on the basis of facts involving ontologically primary entities.

— 449 —

It is this second thesis that mainly drives the naturalist's program. And as views have changed about the kind of explainability that constitutes ontological accommodation, so has the entire program of metaphysical naturalism.

Additionally, the metaphysical viewpoint embodied in (N1) and (N2) has typically been accompanied by a nonskeptical outlook in epistemology, according to which we have access to—and can come to have knowledge of—the phenomena that comprise the natural order. If one combines this nonskeptical outlook with metaphysical naturalism, then one has an overall philosophical package—call it *philosophical*

*naturalism*—that imposes the following metaphysical and epistemological constraints on any nonskeptical version of moral realism:

(C1) If there are any moral properties or facts, they must be naturalistically accommodated.

(C2) If moral knowledge is possible our access to and knowledge of moral properties and facts must be explainable according to epistemological principles we use to explain our knowledge of the natural world generally.

Failure of the moral realist to meet these constraints would evidently have quite a high price tag. The price for non-accommodation is metaphysical queerness of moral properties and facts, plus all the epistemological problems that come with trying to plausibly explain how we have epistemic access to such oddball entities. Certainly these constraints on an acceptable version of moral realism impose quite demanding burdens on the moral realist, but ones well worth shouldering in light of the consequences of failure to do so.

Of course, this particular problematic for metaethics is driven by philosophical naturalism, and there have been those who don't buy that picture and so don't feel constrained by (C1) and (C2). But despite one's global metaphysical and epistemological proclivities, wouldn't it be nice if moral realism could meet these constraints? It would remove most of the metaphysical and epistemological worries about ethics.

### **3. Remembrance of Things Past**

Those who take their philosophical naturalism seriously either have to go the eliminativist route and argue against the presumption that there are moral properties and facts to be somehow naturalistically accommodated, or get down to business and do some accommodating. How might the latter approach be implemented? I.e., if there are indeed objective moral truths, then how might these truths be explainable on the basis of lower-order, non-moral, facts?

Traditionally, this underwriting project was shaped by a certain, quite

— 450 —

narrow, conception of philosophical naturalism—a strongly reductive version which took identity to be the proper relation between successive ontological levels and took some sort of reductive analysis of higher-level terms and concepts to be the only available strategy for securing the needed identities. Furthermore, it certainly appeared—at first, anyway—that the about the only way to approach the reductive task was by way of showing that higher-level terms and expressions had the same meaning as lower-level terms and expressions. So, for instance, a synonymy criterion of property identity was thought to undergird legitimate reductions of one property to another.

The particular reductivist program, applied to moral phenomena, meant that moral properties and facts were to be reductively identified with more basic properties and facts, perhaps psychological or even biological ones. Moreover, the reduction was to proceed by way of analytic meaning connections between moral terms and expressions on the one hand, and nonmoral terms and expressions on the other. In short, moral terms were supposed to have analytic naturalistic definitions, and these were to provide the ultimate basis of the ontological accommodation of the moral to the natural. Call this version of ethical naturalism, *analytic ethical naturalism*.

Of course, if the analytic ethical naturalist could provide the desired meaning reductions, then the two naturalist constraints would be satisfied and moral realism would be in good shape. But we all know what happened: in the early part of this century, G. E. Moore's Open Question Argument lowered the boom on analytical ethical naturalism. Consider any proposed analytic definition of a moral term M by means of a non-moral term or expression N, and consider a corresponding question of the form 'Entity x is N, but is x M?' If the putative definition is correct, then the answer to this question should be knowable simply by reflection on its meaning, and in this sense closed; but invariably, such questions are open.

In light of the Open Question Argument, it appeared that analytic naturalism was simply false. This, in turn, seemed to show that the philosophical naturalist's constraints could not be satisfied *vis a vis* moral properties and facts. So the available alternatives, apparently, were (i) non-naturalist moral realism (Moore's position), or (ii) irrealism about matters moral. Since non-naturalism seemed so philosophically unpalatable, irrealism—in the form of noncognitivist accounts of language—proliferated in the decades following Moore. Moore's Open Question Argument had evidently put ethical naturalism down for the count, and things looked bleak for moral realism.

#### **4. That Was Then, This Is Now**

Times have certainly changed, even in philosophy. A number of recent developments have considerably altered the program of philosophical naturalism, and consequently made possible the re-emergence of a new and

— 451 —

improved version of ethical naturalism. Although the naturalistic outlook still reigns supreme, philosophers have lightened up about the constraints a respectable philosophical naturalism must satisfy. In particular, those wedded to the naturalist program no longer view (N2) as requiring *reduction* of higher-order terms, concepts, theories, or properties to lower-order ones—not in the narrow sense involving inter-level analytic meaning equivalences between terms and expressions of different ontological levels, and also not in the somewhat wider sense involving type/type “bridge laws” involving inter-level *nomic* equivalences. This relaxing of constraints has emerged in the wake of a number of interrelated developments in metaphysics, epistemology, philosophy of language, and philosophy of mind in the 1960's and 1970's.

First, there has been a widespread rejection of a synonymy criterion of property identity, in light of numerous apparent counterexamples. For instance, the (sortal) property *being water* is identical with the property *being composed of H<sub>2</sub>O molecules*; heat is identical with molecular motion; temperature is identical with mean molecular kinetic energy; and so on. But no one supposes that ‘being water’ is synonymous with ‘being composed of H<sub>2</sub>O molecules,’ or that ‘temperature’ is synonymous with ‘mean molecular kinetic energy,’ and so forth for many other scientific identities.

Second, ever since the pioneering work of Kripke (1972) and Putnam (1975), there has been articulation and widespread acceptance of the idea that names and natural kind terms are *rigid designators*; i.e., such expressions designate the same entity with respect to every possible world in which that entity exists. Two important consequences result: (i) identity statements involving rigid designators flanking the identity sign, as in ‘Water = H<sub>2</sub>O,’ are necessarily true without being analytic; and (ii) such statements constitute *definitions*—not the kind that express meaning connections and are thus analytic, but rather *synthetic definitions* that give the real essence of the particular, property, or kind designated by a certain term. Thus, if true, ‘Water = H<sub>2</sub>O’ is a non-analytic necessary truth that expresses the real, underlying, essence of water and thus provides a (synthetically true) definition of ‘water.’

Third, ever since the pioneering work of Putnam (1967a, 1967b), there has been articulation and widespread acceptance of the idea that mental state-types (properties), except perhaps phenomenal properties (“qualia”), are *functional* properties—i.e., their essence consists not in any intrinsic features, but rather in a certain syndrome of *typical causal relations* to other state types, in particular to sensory states, actions, and other mental states which themselves are also functional properties.<sup>1</sup> Functional properties are potentially “physically realizable” quite differently from one occasion of instantiation to another—depending, perhaps, upon the specific physical composition of the cognizer in which they are instantiated.

Fourth, largely because of the influence of functionalism in philosophy

— 452 —

of mind, philosophers have increasingly repudiated their earlier tendency to cash constraint C2 of philosophical naturalism as requiring that the entities (particulars, properties, kinds, facts, etc.) cited by higher-order forms of discourse be *identical* to entities posited by natural science, and ultimately by physics. A more relaxed, California-style, attitude has emerged that counts certain higher-order, multiply realizable, entities as naturalistically respectable—functional properties being the paradigm case. “Naturalness”—i.e., naturalistic respectability—of these higher-order properties and facts consists in (i) *supervenience* upon lower-order properties and facts, and (ii) the *explainability* of specific psychophysical supervenience relations.<sup>2</sup>

Fifth, in recent philosophy of language there has been 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’ uses of a term and the entity 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 by means 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). 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.

Sixth, in epistemology there has been a move away from the idea that there are *a priori* constraints, of the sort associated with the traditional foundationalist program, on acceptable methods of knowledge gathering. Instead, philosophers have begun to move in the direction of "naturalized" epistemology, which recognizes the radical contingency of our methods and procedures of inquiry, especially in science. Accounts of knowledge in terms of such causal notions as the reliability of belief generating processes have begun to displace former views that rested heavily upon *a priori* epistemic principles. Moreover, essentially coherentist holistic frameworks of inquiry have largely displaced the more traditional foundationalist frameworks.

The various philosophical themes and developments just outlined all came together, in a manner highly relevant to the subsequent emergence of new-wave moral realism, in one of the two major species of functionalism in philosophy of mind: so-called *psychofunctionalism*. On this view, mental properties are multiply-realizable functional properties whose relational essences are fully capturable not by the generalizations of common-sense mentalistic psychology ("folk psychology"), but instead by the generalizations of the (ideally complete) *empirical* psychological theory T that happens to be true of humans.<sup>3</sup> According to psychofunctionalism, mental terms

— 453 —

refer rigidly to these properties; and this rigid reference underwrites certain synthetic definitions of mental properties—where the definitive causal role of each such property is specified by means of the empirical theory T.<sup>4</sup> The reason why our mental terms refer to *these* properties is that there are suitable reference-sub-serving causal relations linking (instantiations of) these specific functional properties to people's uses of mental terms and concepts; i.e., the mental terms and concepts "track" the relevant functional properties, in something like the way that radar systems track flying objects, or that traces in a bubble chamber track electrons. Accordingly, it is quite possible to know that certain mental properties are instantiated by certain individuals, even if one does not know the functional essences of the attributed properties, and indeed even if one has numerous false beliefs about those properties. The true nature and essence of mental properties, says the psychofunctionalist, is a thoroughly *empirical* question, to be answered not by *a priori* reflection but by scientific inquiry.

So although Moore's Open Question Argument at first made betting on ethical naturalism a sucker's bet, and turned some form of irrealism into an odds on metaethical favorite, with the emergence on the philosophical scene of an easygoing, nonreductive, naturalism, all bets were off.

## 5. Ethical Naturalism Revived

The various trends noted above, and jointly embodied in psychofunctionalism, are evident in the recent revival of ethical naturalism of the 1980s. To begin with, David Brink, one of most prominent of the new wavers, has suggested that moral properties are functional properties of a certain kind:

[T]he moral realist might claim that moral properties are functional properties. He might claim that what is essential to moral properties is the causal role which they play in the characteristic activities of human organisms. In particular, the realist might claim that moral properties are those which bear upon the maintenance and flourishing of human organisms. Maintenance and flourishing presumably consist in necessary conditions for survival, other needs associated with basic well-being, wants of various sorts, and distinctively human capacities. People, actions, policies, states of affairs, etc. will bear good-making properties just insofar as they contribute to the satisfaction of these needs, wants, and capacities... [and] will bear bad-making moral properties just insofar as they fail to promote or interfere with the satisfaction of these needs, wants, and capacities. The physical states which contribute to or interfere with these needs, wants, and capacities are the physical states upon which, on this functionalist theory, moral properties ultimately supervene (1984: 121–22).

Brink also maintains that moral inquiry is a matter of seeking a normative theory that *coheres* best with both moral and nonmoral beliefs (Brink 1984, 1989). This

— 454 —

coherentist methodology, usually called “reflective equilibrium,” rejects any appeal to *a priori* moral truths or *a priori* constraints on moral inquiry: in ethics, as in science, our methods of knowledge-gathering are radically contingent. For Brink, moral properties are functional properties whose relational essences are captured by whatever specific normative moral theory would emerge, for humans, as the outcome of correctly applied coherentist methodology:

The details of the way in which moral properties supervene upon other natural properties are worked out differently by different moral theories. Determination of which account of moral supervenience is best will depend upon determination of which moral theory provided that best account of all our beliefs, both moral and non-moral (1984: 121, cf. 1989: 175).

One finds the clearest statement of the *semantic* component of new wave moral realism in the work of Richard Boyd (1988), whose position has three key ingredients. First, Boyd proposes to construe moral terms like ‘good’ and ‘right’ (and the concepts they express) as being semantically like natural kind (and other scientific) terms, in having natural, “synthetic” definitions that reveal the essence of the property the term expresses. This means, of course, that moral terms need not have analytic definitions of the sort that were central to more traditional versions of ethical naturalism. Second, the claim that moral terms function this way evidently requires that they are rigid. Like natural kind terms, moral terms allegedly rigidly designate the properties to which they refer. 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 people’s uses of such terms and the relevant natural properties.

According to Boyd's own version of the causal theory of reference, reference is essentially an epistemic notion; so the relevant causal relations 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).

— 455 —

Extending this version of the causal theory 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.

On Boyd's view, then, the fact that humankind's uses of moral terms are regulated in the way described by 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 summarize what we call new wave moral semantics (as developed by Boyd) as the following thesis:

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

A corollary of CSN is that each moral term *t* has a synthetically true natural definition whose definiens characterizes, in purely natural terms, the unique natural property that supposedly casually regulates the use of *t* by humans. (CRT too 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*.)

Boyd's views and Brink's are evidently quite compatible, and in fact are nicely complementary. Brink is explicit in claiming that moral properties are functional properties whose essence is captured by a specific normative moral theory, but says rather little about the semantics of moral terms. Boyd, on the other hand, explicitly claims that moral terms work like natural-kind terms in science and that they designate natural properties, but says rather little about the nature of these properties. So the currently most plausible form of ethical naturalism is an amalgam—the Brink/Boyd view, as we shall call it.<sup>5</sup>



Also an integral part of the Brink/Boyd view is their holistic moral epistemology, involving a coherentist methodology of moral inquiry. In light of CSF, this commitment to epistemological coherentism in ethics is evidently quite compatible with Brink's and Boyd's moral realism. For, if indeed the normative theory T that best coheres with humankind's moral and non-moral beliefs is true, then T will qualify as true not *by virtue* of this coherence—that would be an irrealist, “constructivist,” conception of moral truth—but rather because coherentist methodology is likely, as a matter of *contingent fact*, to converge upon the very normative theory whose generalizations capture the essence of the functional properties that causally regulate the uses of moral terms by humankind, properties that are thus (according to CSF) the referents of moral terms.

This constellation of views—moral functionalism, a holistic moral epistemology, and causal semantic naturalism—together make up what we take to be the most plausible and complete version of new wave moral realism

— 456 —

to date. It is a species of what we call *synthetic ethical naturalism*, to be distinguished from its predecessor, analytic ethical naturalism. And certainly (this version of) synthetic ethical naturalism, with its many similarities to psychofunctionalism in philosophy of mind, has a lot going for it.

Recall the naturalistically inspired metaphysical and epistemological constraints on moral realism. The Brink/Boyd view seems to satisfy them quite well. Take the metaphysical constraint C1, requiring that any moral properties and moral facts must be naturalistically accommodated. According to the Brink/Boyd view, moral properties are perfectly respectable natural properties, viz., multiply realizable functional properties. If so, then contrary to J. L. Mackie (1977), there is nothing metaphysically “queer” about the supervenience relation that obtains between certain lower-order properties and moral properties. The supervenience of moral properties is no more queer than is the supervenience of mental properties, given a functionalist position in philosophy of mind: in either case, specific supervenience connections are naturalistically explainable on the basis of (i) the correct synthetic definitions of the higher-order properties, together with (ii) relevant lower-order naturalistic facts.<sup>6</sup>

The Brink/Boyd view also appears to hold the key for answering an epistemological worry about moral realism, thus satisfying the epistemological constraint C2. In answer to the question, “If there are objective moral properties, what access do we have to them?” the ethical naturalist simply replies that we have access to—and hence come to have knowledge of—such properties in much the way we come to have access to, and come to have knowledge of, other sorts of objective properties. Just as scientific inquiry relies ultimately upon observations (often informed by theoretical assumptions, when it comes to matters highly theoretical in nature), so in moral inquiry we make moral “observations”—viz., spontaneous moral judgments—that provide cognitive access to moral properties and facts, thereby providing the basis for developing an overall coherent system of moral and nonmoral beliefs constituting moral knowledge. Of course, philosophical critics (e.g., Mackie 1977) have frequently argued that facts about disagreement in moral belief, and diversity in people's moral codes, make the idea of moral observations highly suspect. But if people's spontaneous moral judgments are *causally regulated* by the very properties to which

moral terms refer, as new-wave naturalism asserts, then these judgments thereby possess an epistemic status roughly comparable to that of observational judgments in science. I.e., spontaneous moral judgments provide humans with a (defeasible) form of cognitive access to objective moral truth—access that can then become substantially refined, deepened, systematized, and (as need be) corrected by means of coherentist methods of moral inquiry. Although people sometimes disagree in their specific moral judgments, in the moral codes to which they subscribe, and in their conceptions of the properties expressed by moral terms, these facts do not seriously call

— 457 —

into question our access to, or knowledge of, moral properties and facts. For moral terms (as for scientific terms), Boyd's semantic thesis CSN explains (i) how, despite interpersonal variations in moral beliefs (or scientific beliefs), there is indeed some single objective property that is the referent of a given moral (or scientific) term; and (ii) how, by virtue of the causal-regulatory nature of this reference relation, humans have socially coordinated cognitive access to these properties and thus can acquire moral knowledge.<sup>7</sup>

So, the Brink/Boyd version of synthetic ethical naturalism evidently meets very well both the metaphysical and the epistemological constraints imposed by philosophical naturalism. In doing so, this form of moral realism relies fundamentally on a certain semantic view about moral terms. Notice that this fact reveals a parallel between analytic and synthetic ethical naturalism: each species involves a particular moral semantics that plays an absolutely fundamental role in defending the view's naturalistic moral ontology and the associated epistemology. But synthetic ethical naturalism, besides comporting nicely with the anti-reductivist trends in the naturalist's program, also (and as a result) entirely sidesteps Moore's open question argument: who cares if moral terms don't have analytic naturalistic definitions? It appears that new wave moral realism, having entered vigorously into the meta-ethical ring, has everything going for it. But so it seemed with Mike Tyson. The semantic thesis CSN, the linchpin of the Brink/Boyd view, is also its glass jaw.

## 6. Enter: Moral Twin Earth

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., H<sub>2</sub>O) 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 H<sub>2</sub>O as only one of its various actual or physically possible species.) What sort of evidence can be put forth to support the contention that 'water' really does rigidly designate the sortal kind-property H<sub>2</sub>O?

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 H<sub>2</sub>O, 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’

— 458 —

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 English term rigidly designates H<sub>2</sub>O, whereas the twin-English term rigidly designates XYZ; and (ii) that this fact explains why the 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 H<sub>2</sub>O.<sup>8</sup>

The form of argument just canvassed can be called a *semantic competence argument*.<sup>9</sup> Presumably, competent speakers have a comparable intuitive mastery of the semantic workings of ‘good’ and other fundamental moral terms. So if the Brink/Boyd thesis CSN is correct, then things should go the same way they go with ‘water.’ That is, if indeed the term ‘good’ purports to rigidly designate 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; and (ii) these judgments are accompanied by the more general intuitive judgment that ‘good’ does indeed work semantically as CSN says it does. 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 in all probability, CSN is false.

We maintain that things go the latter way—i.e., one’s intuitive judgments concerning a suitable Twin Earth scenario go contrary to CSN. 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 the Brink/Boyd view 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<sup>c</sup>.<sup>10</sup> 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

— 459 —

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 Earthlings 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 ‘good,’ ‘right,’ 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 (as we are already supposing) 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  $T^d$ .<sup>11</sup> 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 twin-moral 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.)<sup>12</sup> In addition, suppose that if Twin Earthlings were to employ

— 460 —

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  $T^c$ , 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  $T^d$ .

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.<sup>13</sup>

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. Yet 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.

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,

— 461 —

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.<sup>14</sup>

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.

The outcome also underwrites the following “open question argument,” a version directed not against analytic semantic naturalism (as was Moore’s original open question argument) but instead against CSN. First premise: If CSN is true, then questions of the following form are *closed*, in the sense that any competent speaker who properly exercises his competence will judge—solely on the basis of his understanding of how the relevant terms work semantically—that the answer to each question is trivially and obviously ‘yes:’

Q1 Given that the use of ‘good’ by humans is causally regulated by natural property N, is entity e, which has N, good?

and

Q2 Given that the use of ‘good’ by humans is causally regulated by natural property N, does entity e, which is good, have N?

Second premise: Questions of the form Q1 and Q2 are *not* closed, as evidenced by one’s intuitions concerning the Moral Twin Earth scenario. Conclusion: CSN is false.<sup>15</sup>

## 7. Down for the Count, Again

Moral Twin Earth packs a mean punch. Moral realism, having returned to the philosophical ring in newly lean and mean naturalistic shape, has been decked again. Furthermore, 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 synthetic naturalism 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

— 462 —

reasoning against CSN should apply, *mutatis mutandis*, against the envisioned alternative version of semantic naturalism.

No doubt various attempts at resuscitation might be made, now that new wave moral realism is flat on the canvass. But in the end, we suspect, all such attempts will prove futile. Moral realism can be squared with the constraints imposed by philosophical naturalism only if CSN, or some similar form of semantic naturalism, is a viable approach to moral semantics. But synthetic semantic naturalism is down for the count.<sup>16</sup>



## References

- N. Block, 1978. "Troubles with Functionalism." In C. Savage, ed., *Minnesota Studies in the Philosophy of Science*, IX. (Minneapolis: University of Minnesota Press.)
- D. Brink, 1984. "Moral Realism and Skeptical Arguments from Disagreement and Queerness," *Australasian Journal of Philosophy* 62.
- D. Brink, 1989. *Moral Realism and the Foundations of Ethics*. (Cambridge: Cambridge University Press).
- R. Boyd, 1988. "How to Be a Moral Realist." In G. Sayre-McCord, ed., *Essays in Moral Realism*. (Ithaca, NY: Cornell University Press).
- T. Burge, 1979. "Individualism and the Mental." In P. French et. al., *Midwest Studies in Philosophy*, Volume 4. (Minneapolis: University of Minnesota Press).
- T. Burge, 1986. "Individualism and Psychology," *Philosophical Review* 95.
- H. Field, 1978. "Mental Representation," *Erkenntnis* 13.
- J. Fodor, 1968. "Propositional Attitudes," *The Monist* 61.
- J. Fodor, 1986. *Psychosemantics*. (Cambridge, MA: Bradford Books/MIT Press).
- T. Horgan, 1984. "Functionalism and Token Physicalism," *Synthese* 59.
- T. Horgan and G. Graham, 1990. "In Defense of Southern Fundamentalism," *Philosophical Studies* 52.
- T. Horgan and M. Timmons, 1991. "Troubles on Moral Twin Earth: Moral Queerness Revived," *Synthese* 66.
- T. Horgan and M. Timmons, forthcoming. "Troubles for New Wave Moral Semantics: The 'Open Question Argument' Revived."
- D. Lewis, 1972. "Psychophysical and Theoretical Identifications," *Australasian Journal of Philosophy* 50.
- B. Loar, 1982. *Mind and Meaning*. (Cambridge, UK: Cambridge University Press).
- W. Lycan, 1981. "Toward a Homuncular Theory of Believing," *Cognition and Brain Theory* 4.
- S. Kripke, 1972. "Naming and Necessity." In D. Davidson and G. Harman, eds., *Semantics of Natural Language*. (Dordrecht: Reidel).

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

— 463 —

H. Putnam, 1967a. “The Mental Life of Some Machines.” In H. Castañeda, ed., *Intentionality, Minds, and Perception*. (Detroit: Wayne State University Press).

H. Putnam, 1967b. “Psychological Predicates.” In W. H. Capitan and D. D. Merrill, eds., *Art, Mind, and Religion*. (Pittsburgh: University of Pittsburgh Press).

H. Putnam, 1975. “The Meaning of ‘Meaning’.” In K. Gunderson, ed., *Minnesota Studies in the Philosophy of Science*, VII. (Minneapolis: University of Minnesota Press).

M. Timmons, 1990. “On the Epistemic Status of Considered Moral Judgments,” *Southern Journal of Philosophy* 29, Spindel Conference Supplement.

S. White, 1982. “Partial Character and the Language of Thought,” *Pacific Philosophical Quarterly* 63.

## Endnotes

<sup>1</sup>Nowadays it is commonly conceded that an adequate functionalist approach must take account of the so-called “wide content” of propositional attitudes, as partially determined by one’s physical and social environment. See Burge (1979, 1986), Loar (1982), White (1982), and Fodor (1986). <sup>2</sup>Philosophers have been rather slow to recognize that supervenience facts must be explainable, rather than *sui generis*, in order to fit smoothly into a broadly naturalistic metaphysics. For elaboration of the point, and discussion of how functionalism provides for the explainability of psychophysical supervenience relations, see Horgan and Timmons (1991). <sup>3</sup>On psychofunctionalism, see Fodor (1968), and Block (1978), Field (1978), and Lycan (1981). The other major species of functionalism, so-called *analytic functionalism*, asserts that the mental properties countenanced by folk psychology are functional properties whose relational essence is capturable by the generalizations of folk psychology itself; cf. Block (1978). <sup>4</sup>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). <sup>5</sup>This position includes, in addition to theses CRT and CSN, the following *functionalist causal regulation thesis*. (Terminology: If T is a normative moral theory and t is a putatively property-expressing moral term, let T(t) be the functional property whose relational essence is just the functional role which, according to the generalizations of T, constitutes the t-role.) FCRT *Functionalist causal regulation thesis*: There is a unique normative moral theory T such that for each moral term t, humankind’s uses of t are causally regulated by the functional property T(t). The theses FCRT and CSN together entail this corollary:— 464 — CSF *Causal semantic functionalism*: Each moral term t rigidly designates the unique functional property that causally regulates the actual-world uses of t by humans. <sup>6</sup>This point is elaborated at some length in Horgan and Timmons



(1991).<sup>7</sup> For further discussion of the points made in this paragraph, with emphasis on the crucial undergirding role played by Boyd's view that moral terms refer to certain natural properties that causally regulate their use by humans, see Timmons (1990).

<sup>8</sup> Some philosophers who espouse causal theories of reference, and who hold that statements like 'Water = H<sub>2</sub>O' 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.<sup>9</sup> 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).<sup>10</sup> Of course, 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 we are here granting this highly optimistic assumption because we want to show that even if it is true, CSN is incorrect anyway. We are also granting for argument's sake that the posited causally regulating properties are indeed *functional* properties whose essence is capturable by some normative moral theory; and this assumption will be built into the specific Twin Earth thought experiment we shall now set forth. But if one were to take a different view about the putative causally regulating properties—say, the view that moral terms (nonrigidly) designate, in a given socio-cultural situation, whatever first-order physical properties (or property clusters) happen to collectively satisfy T<sup>c</sup> (in that situation)—then our Twin Earth story could be modified appropriately. This would not change the moral we shall draw from the story.<sup>11</sup> 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. And since deontological functional properties are reasonably similar, overall, to the kinds of functional properties characterized by consequentialist normative theories, there is also no— 465 — particular reason to doubt that deontological functional properties are *physically realizable* by certain first-order physical properties (or by certain "homeostatic clusters" or first-order physical properties, to use Boyd's terminology). Furthermore, in order for a Twin Earth thought experiment to serve its intended purpose, the relevant scenario need not be a genuine *metaphysical* possibility anyway, but only a broadly *conceptual* possibility. (If Putnam's original Twin Earth scenarios should turn out to be metaphysically impossible, for example, this would not alter their conceptual coherence, and hence would not alter their relevance to the semantics of 'water'.)<sup>12</sup> 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. <sup>13</sup>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 natural properties; (ii) that moral properties are non-natural properties that supervene upon natural ones without being identical to them; or (iii) that moral properties do not exist at all. <sup>14</sup>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. <sup>15</sup>For a considerably more extensive articulation of this form of open question argument, see Horgan and Timmons (forthcoming). <sup>16</sup>This paper is fully collaborative; order of authorship is alphabetical. We benefitted from the discussion of Horgan and Timmons (forthcoming) at the Mark Overvold Memorial Conference, and from Eric Kraemer's commentary on that paper.— 466 —