DAVID BARNETT

AGAINST *A POSTERIORI* MORAL NATURALISM

ABSTRACT. *A posteriori* Moral Naturalism posits *a posteriori* moral/naturalistic identities. Versions of this view that posit necessary identities purport to rely on the Kripke/Putnam doctrine of scientific essentialism. Versions that posit only contingent identities require that moral terms are non-rigid designators. I argue that metaethics does not fall within the scope of scientific essentialism and that moral terms are not non-rigid designators.

INTRODUCTION

A posteriori moral naturalism – or AMN – is given by two claims: that moral properties are identical to natural properties and that such identities are knowable only *a posteriori*. AMN has been endorsed by a significant number of philosophers.[1] I will argue that it lacks justification.

AMN is motivated by two widely held theses: *moral supervenience*, the thesis that two states of affairs cannot, even across all possible worlds, differ morally without differing naturalistically; and *non-conceptual-entailment*, the thesis that there are no purely conceptual moral/naturalistic entailments. By positing moral/naturalistic identities, all versions of AMN comply with *moral supervenience*; and, by positing only *a posteriori* necessary identities, most comply with *non-conceptual-entailment*.[2] AMN is therefore attractive.

The most plausible forms of AMN posit *necessary* identities. They purport to rest on *scientific essentialism*, a doctrine introduced by Saul Kripke and Hilary Putnam which asserts that empirical investigation is needed to discover various necessities, such as that water is H_2O, that Hesperus is Phosphorus, and that heat is the motion of molecules.[3] Scientific essentialism rests on intuitions about hypothetical "twin-earth" scenarios. On the basis of such

Philosophical Studies **107**: 239–257, 2002.
© 2002 *Kluwer Academic Publishers. Printed in the Netherlands.*

scenarios I will argue that forms of AMN which posit necessary identities *cannot* rest on scientific essentialism and therefore lack justification. (In doing so, I hope to at least clarify the role of twin-earth scenarios in metaethical debate.) Near the end of the paper I will argue against forms of AMN that posit only contingent identities.

If my arguments are sound, two options remain for the moral realist: to reject the non-conceptual-entailment thesis and search for purely conceptual moral/naturalistic definitions or to pursue some form of non-naturalism.

1. SCIENTIFIC ESSENTIALISM

Putnam (1975) convinced most of us that a twin-earth stuff satisfying our "intensional" notion of 'water', but having a composition different from samples of water on earth, would not be water. Two conclusions were drawn: that the full meaning of 'water' can be known only *a posteriori*; and that if water is actually composed of H_2O, then it is necessarily composed of H_2O. Kripke (1972) convinced most of us that proper names rigidly designate their referents and therefore, because empirical investigation is sometimes needed to establish coreference, that some identities are necessary *a posteriori*. Kripke and others then generalized this result over natural-kind terms. They concluded for instance that 'water' rigidly designates H_2O. If 'H_2O' and 'water' rigidly co-refer, then it is necessary both that water = H_2O and that being water = being H_2O.

The aim of this section is to clarify the scientific-essentialist – or SE – method for establishing *a posteriori* necessary property-identities.

The scientific essentialist begins by considering a property F and a property G which are not *a priori* coextensive. Hypothetical scenarios are then used to elicit intuitions as to whether F or G might be instantiated without the other. Intuitions in favor of such a possibility are *anti-SE* intuitions; those against it are *pro-SE* intuitions. Pro-SE intuitions, in the absence of anti-SE intuitions, are taken to signal an *a posteriori* necessary identity. The presence of both pro- and anti-SE intuitions is taken to signal the need for further adjudication.

Putnam's twin-earth scenario elicits a pro-SE intuition: if scientists have not been deceived about the composition of water – if all water on earth is composed of H_2O – then, intuitively, a stuff macroscopically identical to water but composed of XYZ ($\neq H_2O$) on a superficially similar planet would not be water. This intuition is (or at least *was*) however accompanied by anti-SE intuitions, such as that scientists might have discovered that water was composed of something other than H_2O, or that a puddle of water might contain no hydrogen.

Kripke suggests that when we appear to have both pro- and anti-SE intuitions our intuitions are actually in a state of accord, though we are misreporting a mere epistemic possibility as metaphysical one. He asserts that a rephrasal strategy is needed to deflate such erroneous reports. George Bealer (1994, pp. 193–198) effectively argues that Kripke's *Naming and Necessity* contains two rephrasal strategies; that the first succeeds; and that the second fails due to its lack of asymmetry – that is, SE-opponents can use it equally well to disarm *pro*-SE intuitions. Kripke gives his first rephrasal strategy as follows:

> [I]n advance, we are inclined to say, the answer to the question whether Hesperus is Phosphorus might have turned out either way ... There certainly is a possible world in which a man should have seen a certain star at a certain position in the evening and called it 'Hesperus' and a certain star in the morning and called it 'Phosphorus'; and should have concluded – should have found out by empirical investigation – that he names two different stars ... [A]nd so it's true that given the evidence that someone has antecedent to his empirical investigation, he can be placed in a sense in exactly the same situation, that is a qualitatively identical epistemic situation, and call two heavenly bodies 'Hesperus' and 'Phosphorus', without their being identical. (1972, pp. 103–104).

Here, Bealer charitably generalizes Kripke's proposal and explains why it meets the asymmetry criterion:

> ... to deflate the force of an intuition reported with ⌜Possibly A⌝, one rephrases that report with ⌜It is possible that a population of speakers in an epistemic situation qualitatively identical to ours would make a true statement by asserting ⌜A⌝ with normal literal intent⌝. Because our anti-scientific-essentialists are traditional internalists, they are committed to holding that the meaning of ⌜A⌝ cannot differ across populations of speakers in qualitatively identical epistemic situations. Accordingly, they must hold that the rephrasal entails the original report. In particular, when the original intuition seems to have a pro-SE force, they are committed

to holding that the rephrasal has that force as well. By contrast, scientific essen-
tialists are not traditional internalists, so they are free to hold that the meaning of
$\ulcorner A \urcorner$ can differ across populations of speakers in qualitatively identical epistemic
situations. So when the original intuition seems to have an anti-SE force, they are
free to hold that that force is deflated upon rephrasal (1994, pp. 197–198).

Thus, in reporting that a puddle of water might contain no hydrogen
atoms, we allegedly meant that speakers in an epistemic situation
qualitatively identical to ours might utter a truth by asserting, "a
puddle of water might contain no hydrogen atoms," with literal
intent.[4] In such a situation 'water' would not refer to water, and
'hydrogen atoms' might not refer to hydrogen atoms. The other
expressions in the report are *semantically stable* – they cannot
vary in meaning across populations in epistemic situations qualita-
tively identical to ours.[5] The rephrasal deflates our anti-SE report
by implicating as its source an intuition about mere epistemic
possibility.

   Note that while twin-earth scenarios function primarily to elicit
pro-SE intuitions, they can, if appropriately designed, also aid in
the adjudication stage of the SE-procedure. To play this dual role,
a scenario must allow us to consider both whether F or G might
be instantiated without the other and whether a population in an
epistemic situation qualitatively identical to ours might mean some-
thing different by $\ulcorner F \urcorner$ or $\ulcorner G \urcorner$. To do this, the twin-earthlings'
epistemic situation must be qualitatively identical to ours, so that
the naturalistic facts at the level of cross-earth divergence are
guaranteed to be irrelevant to competence in moral vocabulary.

   In sum, to establish an *a posteriori* necessary identity between
F and G, it must be shown (a) that F and G are coextensive; (b)
that pro-SE intuitions are present; and (c) that anti-SE intuitions,
if present, are deflatable. Since instances of (a) cannot be known *a
priori*, they are for philosophical purposes conditionally assumed to
see whether (b) and (c), the philosophically interesting propositions,
hold.[6] To establish (b) and (c) with respect to a certain domain of
properties is to establish that that domain falls within the scope of
SE.

2. AGAINST A *POSTERIORI* MORAL NATURALISM

First I will argue against a specific SE-form of AMN. Then I will generalize this argument to undermine all SE-forms of AMN. Last, I will argue against forms of AMN that do not purport to rest on SE.

As a preliminary I will introduce *moral functionalism*, which provides for a taxonomy of most species of AMN. Moral function-alism is modeled after functionalist stances in philosophy of the mind. According to it, a nexus of general and specific constraints governs the use of moral language. Of the general sort, there is the principle of moral supervenience. Of the specific sort, associated with each moral predicate is a family of constraints taken to reflect necessary relations among moral properties – e.g., a morally right action is always fair – and among moral and natural properties – e.g., it is generally wrong to inflict gratuitous pain on others.

Jackson and Pettit (1995) provide a nice explication of moral functionalism together with a provisional outline of fragments of our folk moral-functional theory.[7] According to them, a successful folk moral theory is a systematization of all, or a majority of, the *a priori* commonplaces associated with each moral term or expression. They list, for example, seven commonplaces illustrative of the correct use of the term 'fairness'. Here is one of them:

In most contexts 'I cut, you choose' is a fair sort of procedure: if anything is fair, then that is. And similar exemplary status is enjoyed by equal division and by any procedure that departs from equal division only so far as independent claims require. Other arrangements are judged to be fair on the basis of similarity with such paradigms (1995, pp. 22–23).

The list continues, for all moral predicates. Ideally, the folk theory will converge on a stable set of accepted commonplaces, which will serve to define moral-functional properties – properties defined by their locations in the resultant nexus of moral commonplaces.

Moral-functional properties are given formally by the method of *Ramsification*: moral terms are replaced in the moral-functional theory with variables 'N_1', 'N_2', ..., $\ulcorner N_n \urcorner$, which are bound by existential quantifiers ranging over first-order natural properties.[8] Naturalistic application-conditions for moral predicates then fall out in the form of *Ramsey sentences*: e.g., x is right iff there exist first-order natural properties N_1–N_n satisfying our moral-functional

theory and x has N_m (where $\ulcorner N_m \urcorner$ has been substituted for 'right' in our theory). The first-order natural property N_m is then said to *realize* the second-order functional property *having a property that plays the rightness-role defined by our finished theory*. It is an empirical matter to identify which, if any, first-order properties are in fact playing the relevant functional roles.

Species of moral functionalism can be singled out with the aid of the following questions: (i) Are moral terms rigid designators? (ii) Are moral properties identical to second-order functional properties, or to first-order naturalistic realizer properties? (iii) Is the correct moral-functional theory knowable *a priori*?

While nonfunctionalistic forms of AMN are available, they are initially less plausible, for it is unlikely that moral terms are definable (in naturalistic terms) independently of one another in a noncircular fashion. Functionalistic stances acknowledge the interdependence of moral notions while avoiding the threat of circularity by employing the method of Ramsification. Nevertheless, my argument against AMN will not depend on a functionalistic interpretation of it.

My argument begins with a twin-earth based attack on a specific SE-form of AMN. Jackson and Pettit consider the possibility that traditional *a priori* methods will lead us to our finished moral-functional theory, which in turn will pick out a set of naturalistic realizer properties that (i) happen to fill the relevant functional roles defined by our theory, (ii) can be identified only empirically, and (iii) are rigidly designated by our moral terms (1995, p. 28; also Jackson 1998, p. 144).

Suppose, then, that *a priori* reasoning leads us to conclude that a Rawlsian theory of justice containing the general principles for determining the proper distribution of goods and burdens in a society is definitive of our moral-functional theory. That is to say, suppose that our moral-functional theory converges, by way of traditional *a priori* investigation, onto a theory whose content, at the level of functional-roles, is captured by a Rawlsian theory of justice. (I should emphasize that my argument uses the Rawlsian theory only for purposes of illustration: I could have used Mill's theory, or possibly Aristotle's theory, or perhaps even Kant's theory – there are some delicate interpretive matters here.)

Crucial to functionalist versions of AMN is the role played by normative moral theory. A normative theory may characterize the moral-functional theory; that is, it may be definitive of the moral-functional roles associated with our moral language. Or, it may characterize the set of natural properties that fill these roles; that is, it may identify those natural properties that contingently realize the moral-functional theory. The theory definitive of the moral-functional roles may or may not be the same theory that characterizes the realizers of these roles – this is left for the scientists to decide. For instance, happiness utilitarianism may characterize our moral-functional theory while some other theory, say T^n, might turn out to tell us that on earth natural property G maximizes creatures' happiness and therefore serves to fill the rightness-role. That is, it may turn out (i) that the second-order functional property of *having some natural property which plays the rightness-role* is identical to the second-order property of *having some natural property which plays the happiness-maximizing role* and (ii) that G contingently plays the happiness-maximizing role.

Return now to the supposition that a Rawlsian theory of justice, T^J, characterizes our moral-functional theory. Let us further suppose that our economists, guided by the set of general principles embodied in our Rawlsian theory, have empirically discovered that normative theory T_1 identifies the set of natural properties that fill the roles of our moral-functional theory. T_1 tells us in practical terms how we ought morally to distribute the goods and burdens on earth. For each moral property F, T_1 provides a natural property G that has been empirically determined to be coextensive on earth with F. The view under consideration claims additionally that, for each F and for each corresponding G, F and G are *necessarily* coextensive. Justification for this further claim requires a trip to twin-earth.

Our twin-earth scenario should presuppose only that scientists have discovered that F and G are coextensive on earth. It should aid us in considering whether F or G might be instantiated without the other. If it is to serve a dual role – if it is to be used both to elicit pro-SE intuitions and to aid in the rephrasal of any anti-SE intuition reports – the epistemic situation of the twin-earthlings should be qualitatively identical to our own.

Consider "moral twin-earth," TE. TE is at first glance indistinguishable from earth. The macroscopic qualitative features of TE correspond perfectly to those of earth. Furthermore, the epistemic situation of the twin-earthlings is – at least until the time at which our economists discover that T_1 identifies the realizers of T^J on earth – qualitatively identical to ours. When Alice feels guilty, her doppelgänger on TE feels guilty. When Fred looks out at the ocean, Fred's doppelgänger looks out at the "ocean." What it is like for Fred to look out at the ocean is what it is like for Fred's doppelgänger to look out at the "ocean."

Suppose that while visiting TE our economists discover that normative theory T_2 ($\neq T_1$) identifies the natural properties that fill the roles of T^J on TE. T_2 differs from T_1 as a result of any number of subtle cross-earth differences in accidental features in environment. There might for instance be subtle cross-earth genetic differences (assuming our doppelgängers have genes); or the "air" might be uniformly thicker on twin-earth, thereby slowing the transportation of goods there. (While such differences may seem far-fetched, we must keep in mind just how radical the view under evaluation is: according to it, the nature of moral properties is tied up with *hidden accidental features of our environment*.) Under the prescribed principles set forth in our Rawlsian theory, which remains fixed across earths, T_2 identifies the naturalistic realizers of T^J and thereby guides the morally correct distribution of goods and burdens on twin-earth.

Now we survey our intuitions. Given (i) that we have arrived via traditional *a priori* methods at T^J (a moral-functional theory whose content is captured by a Rawlsian theory of justice) and (ii) that T_1 is the normative theory which identifies the realizers of T^J on earth (all and only just acts on earth conform to T_1), we ask ourselves, would it be just to distribute the goods on twin-earth according to T_2, which correctly identifies the realizers of T^J on twin-earth but is not identical to T_1? That is, given that all and only just acts on earth are empirically known to accord with one set of naturalistically described instructions for action, would a twin-earth action in discord with these instructions, but having a natural property that satisfies the justice-role of our *a priori* moral-functional theory, be just?

Intuitively, it would. Pro-SE intuitions are absent. Given that moral properties are coextensive with certain natural properties on earth, we have the intuition that they could have been coextensive with different ones. That is to say, not only are pro-SE intuitions lacking, they are actually replaced by concrete anti-SE intuitions. Of course, the presence of anti-SE intuitions in the absence of pro-SE intuitions spells failure for the view under consideration.

A few comments on what this intuitive evidence does *not* establish are in order. First, it does not establish that moral properties are *multiply realizable* by natural properties, or that moral properties can be *constituted* of different natural properties across possible worlds. These conclusions carry substantial commitments about the relation between moral and natural properties – commitments not warranted by the results of our scenario. Keep in mind that it was only for the sake of argument that we supposed morality to have a functionalistic characterization in the first place. Second, it does not establish that for any natural property N and any moral property M it is possible for M to be instantiated in the absence of N. This would rule out *a priori* moral naturalism. Our scenario restricted the range of considered natural properties with specific suppositions regarding both conceptual and empirical connections among the relevant moral and natural properties. Most importantly, the range was restricted to natural properties that are not *a priori* identifiable with moral properties. So the force of our anti-SE intuition reported above is not threatened by the possibility of our finding it intuitive that if rightness is coextensive with, say, minimizing suffering, then it is necessarily so, for there is plausibly some *a priori* link between minimizing suffering and rightness.[9]

Since there is no stalemate, there is no need to implement our rephrasal strategy, which serves as an adjudicatory device. Nevertheless, the rephrasal strategy would, if implemented, fail to deflate anti-SE intuitions. This result follows intuitively from TE. Before any serious science was done on earth or twin-earth, the twin-earthlings were by stipulation a population of speakers in an epistemic situation qualitatively identical to our own. And, intuitively, twin-earthlings and earthlings meant the same by statements such as, "It is just to distribute goods and burdens according to the principles embodied in T^J." Since the rephrasal strategy depends for

its success on a semantic *instability* in the relevant expressions, it would not in this case have been able to deflate anti-SE intuitions, had there been any.

I conclude that the SE-form of AMN under consideration is unjustified. It appeals to SE but fails both steps of the standard procedure for establishing SE-status.

Before broadening the scope of my argument to include all SE-forms of AMN, I should stress the initial implausibility of a certain class of such positions. The class I have in mind is united by the claim that all moral theory is to be developed at *all* stages by moral scientists – contra armchair philosophers – in the spirit of naturalized epistemology.

On these views, all areas of normative ethics are to be "natural-ized" in a similar fashion to Quine's naturalized epistemology. It strikes me as implausible that a *normative* discipline must turn to science for answers to its fundamental questions: ethics, on the face of it, is concerned with how we *ought* to act, not how we in fact act, how we would act under any naturalistically described circum-stances, or how we would agree to act under any naturalistically described circumstances.

I should emphasize the difference between these doctrines and the SE-form of moral functionalism argued against above. In the case of the latter, the motivation for positing *a posteriori* moral/naturalistic identities derives, at least in part, from tradi-tional *a priori* methods of theoretical ethics. On this view, *a priori* theorizing results in a moral-functional theory. In this context the question naturally arises whether moral properties are identical to functional or to first-order realizer properties. By contrast, in the case of the doctrines under consideration, the declaration that ethics is SE-territory has absolutely no *a priori* motivation from theoretical ethics itself. For according to these doctrines moral theory contains no *a priori* elements. Of course, any challenge to the intuitive, tradi-tional distinction between the methodology of ethics and that of natural science must be either internal or external to ethics.

Suppose that it is internal. That is, suppose that it is alleged that theoretical ethics itself offers reason to doubt the need for *a priori* investigation of its central issues. In order to justify a depar-ture from the standard epistemic norms of a discipline from *within*

that discipline, it is necessary to appeal to the standard norms of that discipline. In the case of theoretical ethics, the standard norms have significant *a priori* elements. We should expect, then, that to justify a departure from this standard from within theoretical ethics would require the employment of at least some *a priori* reasoning. However, according to the positions at hand, there are absolutely no *a priori* elements to be found in ethics. It is not plausible, then, that justification for these radical positions lies internal to ethics.

Suppose, then, that it is external – that these views are justified on purely metaethical grounds. What could such grounds be? While proponents of the views are likely to be motivated by general epistemological and metaphysical doctrines – e.g., methodological or metaphysical naturalism – they can hardly look to such doctrines for justification. That would be question-begging.[10] The only place they can look for justification is to SE. But, as we shall now see, the entire domain of metaethics falls outside the scope of SE.

It is unclear what would be the exact content of the appropriately adjusted twin-earth scenarios for each SE-form of AMN. As already noted, we cannot simply quantify over all natural properties and consider whether, say, rightness, might be instantiated on twin-earth in the absence of any one of them. First, we would not likely have concrete intuitions about such an abstract issue. Second, reports of pro-SE intuitions would not be reliable; for, under such conditions, they might simply be indicative of *a priori* moral/naturalistic identities.

Fortunately, we can avoid the task of designing particular twin-earth scenarios to fit each of the remaining SE-forms of AMN. For we have good reason to judge that pro-SE intuitions will be absent with respect to *any* appropriately designed relevant twin-earth scenario. We simply need to focus our attention on the sorts of natural properties that (i) may be found around here, on earth, and (ii) are hidden from *a priori* investigation. Suppose that it were empirically determined that one of *these* natural properties were coextensive with rightness. Would lacking such a property necessarily disqualify an action on some distant planet, or in some other possible world, from being right? Intuitively, the answer is no. Pro-SE intuitions are absent regarding the general claim that there are necessary *a posteriori* moral/naturalistic identities.

As a consequence, all doctrines that posit *a posteriori* necessary moral/naturalistic identities are destined to fail the first step of the SE-qualifying procedure.

Furthermore, it follows directly from our original case that all such positions are destined also to fail the second step. We concluded that the meaning of 'It is just to distribute goods and burdens according to Rawlsian principles' cannot vary across populations of speakers in qualitatively identical epistemic situations. The upshot of our original rephrasal attempt is that moral expressions are *semantically stable*: their meanings cannot vary across populations of speakers in epistemic situations qualitatively identical to our own. This result entails that the rephrasal strategy will fail for *any* theory that posits moral/naturalistic *a posteriori* identities. If an area of inquiry is currently justified as falling within the scope of scientific essentialism, its characteristic expressions will be semantically *unstable*, like 'water', 'tiger', 'gold', and 'C-fiber'.

I conclude that all forms of AMN which appeal to SE are unjustified, for they are destined to fail both steps of the SE-qualifying procedure.[11]

Before examining non-SE forms of AMN, I need to consider an objection to my argument so far. A methodological naturalist might reject my implicit premise that philosophical intuitions are a legitimate source of evidence. It is, however, hard to see how SE-forms of AMN could themselves be justified without the use of intuitions; clearly, the standard arguments for SE make heavy use of philosophical intuitions.

Still, it might be thought that an alternative justification for SE-forms of AMN lies elsewhere, say, in legitimate moral explanations of naturalistic phenomena.[12] This sort of justification would comply with methodological naturalism. However, it has two problems.

First, the mere existence of such explanations would not justify AMN (whether of an SE-form or not). For such explanations are perfectly consistent with the truth of *a priori* moral naturalism. That is, even if we suppose (i) that there are legitimate moral explanations of naturalistic phenomena, and (ii) that (i) is explained by the fact that moral properties *are* naturalistic properties, it is still open that moral/naturalistic identities are knowable *a priori*.

Second, whatever explanatory value there might be in positing moral/naturalistic connections, it cannot support the strong modal claims of SE-forms of AMN. To see the difficulty here, consider the water/H_2O case. Philosophers have never denied the explanatory value of the hypothesis that water is H_2O: H_2O has just the right properties to explain the superficial and deep properties of water. However, before the advent of SE it was generally accepted that the purview of science was limited to the realm of the contingent. Accordingly, it was thought that it was a contingent fact that water was H_2O.

SE makes the strong modal claim that if water is H_2O, then *necessarily* water is H_2O. The justification for this claim cannot derive solely from its empirical explanatory value. For whatever such value is had by the hypothesis *that necessarily water is H_2O* is already exhausted by the weaker hypothesis *that water is H_2O*. The strength of SE-claims outruns the possible strength of justifications based solely on empirical explanatory value.

Quine, the leader of methodological naturalism, recognized that strong modal claims cannot rest on purely empirical grounds but saw this as no obstacle for his own views, which evidently reject all strong modal claims (and certainly reject *a posteriori* strong modal claims). Some of Quine's followers, on the contrary, seem to embrace SE without fully appreciating its methodological commitments. They seem to ignore that the only available justification for SE makes heavy use of intuitions.

The methodological naturalist moral realist can retreat by positing only contingent *a posteriori* moral/naturalistic identities. However, I shall now argue against such views.

My argument will focus on a form of moral functionalism which I take to offer the most plausible account of "non-rigid" AMN. According to it, (i) *a priori* methods will lead us to the correct moral-functional theory and (ii) moral terms *non*-rigidly designate whatever naturalistic properties happen to fill the relevant functional roles. This version of AMN conflicts with the non-conceptual-entailment thesis. Still, it offers an apparent avenue of escape for AMN theorists not concerned with the thesis.

On this view, rightness $=_{df}$ the first-order property that fills the rightness-role (of our moral-functional theory T). From this defi-

nition together with the plausible thesis that the property which fills the rightness-role might not have been the property which fills the rightness role (just as the tallest mountain might not have been the tallest mountain) it follows that rightness might not have been rightness.[13]

Now, if we think of 'rightness' as a disguised definite description, we might convince ourselves that there is a sense in which rightness might not have been rightness. Theoretical views aside, however, it is strongly counterintuitive that rightness might not have been rightness. Intuitively, rightness has always been, will always be, and *must be* nothing more and nothing less than *rightness*. Let us not repeat Euthyphro's mistake by confusing right actions with rightness itself. It might be contingent that certain right actions are right, even though it is necessary that rightness is rightness. Furthermore, what makes some action right might, at some level, differ from what makes some other action right. This, too, has no bearing on whether rightness itself might have been something other than rightness.

That there is a sense in which it is counterintuitive that rightness might not have been rightness is not, by itself, conclusive evidence against non-rigid forms of AMN. For if 'rightness' is a disguised definite description, scope ambiguity might give rise to several readings of 'rightness might not have been rightness', one of which would be false. This would explain the sense in which it is counterintuitive that rightness might not have been rightness. There are however two problems with this explanation. First, the sentences 'rightness might not have been rightness' and 'the tallest mountain might not have been the tallest mountain' differ in that the former, but not the latter, has a clear and obvious sense in which it seems false. Second, theoretical prejudices aside, there does not seem to be *any* sense in which 'rightness might not have been rightness' seems true or even permits of an alternative reading.

This is not a theoretical argument, but simply an appeal to intuition (which is not to say that it is inconclusive). In an analogous vein, 'squareness', 'bachelorhood', 'equality', 'correctness', and 'shyness' are, intuitively, rigid designators. Might shyness have been anything other than shyness? The intuition becomes even stronger when we consider gerundive formulations. It seems for instance absurd that the property of being right might not have

been the property of being right; surely 'the property of being right' rigidly designates the property of being right.

But 'rightness' and 'the property of being right' are simply nominalizations of the predicate 'is right' – they are property-abstract formulations referring to the property expressed by 'is right'. What we are doing, then, when we Ramsify in terms of the predicate 'is right' is providing a definition of *the property of being right*. If 'the property of being right' is a rigid designator, we have a reason to reject the claim that the property of being right $=_{df}$ the first-order property that fills the rightness role (because the definite description on the right-hand side of the identity sign is not a rigid designator). And if we reject this definition, then we ought to reject the claim that in Ramsifying in terms of 'is right' we are providing a definition of 'the first-order property that fills the rightness role'. What we are defining is *the property of being right* and clearly not a first-order realization of the rightness-role.

The functionalist form of AMN under consideration, then, has counterintuitive consequences and should be rejected. These consequences, furthermore, are consequences of all non-rigid forms of AMN. We therefore have good reason to reject all forms of AMN that avoid appeal to SE, since all entail that moral terms are non-rigid designators.[14,15]

I conclude that all forms of AMN are untenable. If this is right, the moral realist must either reject the non-conceptual-entailment thesis and search for purely conceptual moral/naturalistic definitions or pursue some form of non-naturalism.

NOTES

[1] For instance, Richard Boyd (1988), David Brink (1984), Frank Jackson and Philip Pettit (1995), Frank Jackson (1998), Peter Railton (1993), and Nicholas Sturgeon (1988).

[2] All versions of AMN that posit only necessary identities, and some that posit contingent identities, comport with *non-conceptual-entailment*, in the following sense: they posit no *a priori* strict conditionals with naturalistic antecedents and moral consequents. One version of AMN that posits only necessary identities does, however, posit *a priori* material conditionals with naturalistic antecedents and moral consequents (for an example of this sort of view, see Jackson and Pettit (1995) and Jackson (1998)).

[3] These examples involve names, mass terms, common nouns, and descriptions; neither Kripke nor Putnam has argued explicitly for *a posteriori* identities involving gerundive or property-abstract phrases – syntactical devices used to denote properties. Since our subject matter involves properties, I will hereafter paraphrase Kripke and Putnam using gerundive phrases and property-abstract formulations.

[4] Intuitively, this rephrasal captures the truth conditions, but not the exact content, of what was meant by the original report. This is enough to deflate the anti-SE force of the original report.

[5] The expression 'semantically stable' was introduced by Bealer (1994).

[6] It might be thought that twin-earth arguments must suppose not only that F is coextensive with G, but that an *explanatory* relation holds between F and G. But pretend (i) that being poison and being strychnine are coextensive on earth *and* (ii) that being strychnine *explains* why samples of poisonous substances are harmful when ingested. To invalidate the absurd SE-hypothesis that being poisonous = being strychnine we simply consider a twin-earth where a stuff superficially like strychnine – clear, odorless, causes death when ingested, etc. – has a different composition. Would *that* stuff be poison? Of course it would. Our pretense that being strychnine explains certain features of all poisonous substances on earth is not relevant to the outcome of our twin-earth scenario, which delivers anti-SE results whether or not we take it into consideration. The same can be said in cases where pro-SE results are delivered: it is intuitive to me that if atoms are made of quarks, atomish things made of tworks ($\neq$quarks) would not be atoms, even though I do not know how being made of quarks explains the behavior of atoms.

[7] See also Jackson 1998, Chapter 5.

[8] Note that by restricting the range of the quantifiers in this way, non-naturalistic moral realist positions are excluded. The non-naturalist, then, might endorse the finished moral-functional theory but maintain that the quantifiers introduced during Ramsification range over irreducibly non-natural moral properties and that these are the only properties that can play the relevant roles.

[9] On a related point, one might worry that my scenario does not identify the natural properties that actually fill the functional roles. However, the particular theory that accurately captures the relevant set of natural properties is irrelevant

to our intuitions regarding twin-earth (there might not even be such a theory). In regard to the water/H_2O case, the same pro-SE intuitions would have resulted if Putnam had put his case as follows: if all and only samples of water on earth are actually composed of XYZ ($\neq H_2O$), and if the watery stuff on twin-earth were composed of UVW ($\neq$XYZ), then would UVW be water? Again, our intuitions tell us that UVW, which would not be composed of the *actual* molecular constituents of water, would not be water. What is important is not the particular content of the considered identity, but that it be *a posteriori*. In the water/H_2O case the proposed identity is obviously *a posteriori*. In the moral/naturalistic case more care must be taken to ensure that the considered identity is not knowable *a priori*.

[10] See Railton (1989, pp. 159–160) for a discussion of this point.

[11] On the basis of a twin-earth scenario, Horgan and Timmons (1991, 1992) argue that moral realism lacks the resources to explain *moral supervenience* because (1992, p. 248) "there just do not exist enough semantic constraints on moral terms and concepts for these constraints, together with objective nonmoral facts, to determinately fix the extensions (at a possible world) of moral terms and concepts." Their twin-earth, which I shall call 'TE*', differs from earth in two respects:

(R)   use of the expression 'right' is "regulated" on twin-earth by a natural property distinct from the one that "regulates" it on earth;

and

(P)   our doppelgängers have psychological temperaments slightly different from our own.

It is further stipulated that, due to (P), twin-earthlings would apply their predicate 'is right', whereas we would not, to a certain type of action. Horgan and Timmons claim that it is nevertheless intuitive that our doppelgängers would mean what we do by 'right' and therefore that the semantic constraints on moral language underdetermine (putative) relations between nonmoral and (putative) moral facts.

Their argument seems to me to have at least two problems. The first has to do with (R), whose components are as follows:

(i)   Twin-earthlings' use of 'right' is "regulated" by a natural property, say, N.

(ii)   Earthlings' use of 'right' is "regulated" by a natural property, say, N*.

(iii)   N$\neq$N*.

Horgan and Timmons cite Boyd (1988, p. 195) as the source of their technical term 'regulates'. Boyd's definition includes the following stipulation:

(iv)   k *regulates* term t only if there is ". . . social transmission of . . . beliefs regarding k, formulated as claims about t . . ."

From (i), (ii) and (iv), then, we can infer (v) and (vi):

(v)   A sincere twin-earthling utterance of ⌜a is right⌝ expresses a belief about N, in the sense that it is true iff N is instantiated by a.

(vi)   A sincere earthling utterance of ⌜a is right⌝ expresses a belief about
        N*, in the sense that it is true iff N* is instantiated by a.

It is clear from (iii), (v), and (vi) that earthling and twin-earthling utterances of
⌜a is right⌝ have distinct truth-conditions. They therefore have distinct meanings:
twin-earthlings do *not* mean what we do by 'right'. There is no room then for
intuitions about whether twin-earthlings mean what we do by 'right'. The question
is closed by the stipulations of TE*, and the answer is no.

    The second problem concerns (P). How are we expected to have strong
intuitions about what is meant by a population whose epistemic situation is
qualitatively unlike our own? (R) aside, it seems possible both for the twin-earth-
lings to mean what we do by 'right' and for them to mean something different.
The stipulations of TE* are too strong with (R), and too weak without it.

[12] See, for example, Sturgeon (1988) and Railton (1989).

[13] This view is reminiscent of a physicalist view of the mind put forth by David
Lewis (1980). Lewis claimed that 'pain' was a non-rigid designator and there-
fore that "Pain might not have been pain" (p. 218). Sydney Shoemaker (1981)
responded with a cogent rebuttal, convincing many that "non-rigid" approaches
in philosophy of mind are untenable.

[14] It might be suggested that the counterintuitive consequences of such views
can be explained away by appeal to a theoretical account of contingent identity
along Lewis-style lines. This, however, would miss the place of modal intuitions
in the present debate. Of course when Kripke and Putnam first argued for SE, their
opponents, descriptivists, knew how to respond *as descriptivists*: XYZ *is* water.
Fortunately, sophisticated descriptivists put aside their theory and leant an ear to
their naive intuitions. After conceding the existence of such intuitions it would be
inappropriate to add, "the counterintuitive consequences of descriptivism might
be explained away by appeal to an account of natural-kind terms along descrip-
tivist lines." Likewise, the counterintuitive consequences of an application of
Lewis's non-rigid approach to moral terms cannot be explained away by appealing
to Lewis's own theoretical apparatus.

[15] Note that the arguments of this section do not presuppose any theory of refer-
ence for moral terms, and so they cannot be avoided simply by turning to a new
theory of reference.

REFERENCES

Bealer, G. (1994): 'Mental Properties', *The Journal of Philosophy* 91, 185–208.

Bealer, G. (1998): 'Self-Consciousness', *Philosophical Review* 106, 69–117.

Boyd, R. (1988): 'How to Be a Moral Realist', in G. Sayre-McCord (ed.), *Essays
    in Moral Realism* (pp. 181–182), Ithaca: Cornell University Press.

Brink, D. (1984): 'Moral Realism and Skeptical Arguments from Disagreement
    and Queerness', *Australasian Journal of Philosophy* 62, 111–125.

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

Darwall, S., Gibbard, A. and Railton, P. (1992): 'Toward *Fin de siècle* Ethics: Some Trends', *Philosophical Review* 101, 115–189.

Donnellan, K. (1983): 'Kripke and Putnam on Natural Kind Terms', in C. Ginet and S. Shoemaker (eds.), *Knowledge and Mind, Philosophical Essays* (pp. 84–104), New York: Oxford Press.

Horgan, T. and Timmons, M. (1991): 'New Wave Moral Realism Meets Moral Twin Earth', *Journal of Philosophical Research* XVI, 447–465.

Horgan, T. and Timmons, M. (1992): 'Troubles on Moral Twin Earth: Moral Queerness Revived', *Synthese* 92, 221–260.

Jackson, F. (1998): *From Metaphysics to Ethics, A Defense of Conceptual Analysis*, New York: Oxford University Press.

Jackson F. and Pettit, P. (1988): 'Functionalism and Broad Content', *Mind* XCVII, 381–400.

Jackson F. and Pettit, P. (1995): 'Moral Functionalism and Moral Motivation', *The Philosophical Quarterly* 45, 20–39.

Kripke, S. (1980): *Naming and Necessity*, Cambridge: Harvard University Press.

Lewis, D. (1980): 'Mad Pain and Martian Pain', in N. Block (ed.), *Readings in the Philosophy of Psychology*, Vol. 1 (pp. 216–222), Cambridge: Harvard University Press.

Lewis, D. (1995): 'Lewis, David: Reduction of Mind', in S. Guttenplan (ed.), *A Companion to the Philosophy of Mind* (pp. 412–431), Cambridge: Basil Blackwell Publishers.

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

Moore, G.E. (1922): 'The Concept of Intrinsic Value', in his *Philosophical Studies* (pp. 253–275), New York: Harcourt, Brace, and Co.

Putnam, H. (1975): 'The Meaning of "Meaning"', in K. Gunderson (ed.), *Language, Mind, and Knowledge: Minnesota Studies in the Philosophy of Science*, vol. 7 (pp. 131–193), Minneapolis: University of Minnesota Press.

Railton, P. (1989): 'Naturalism and Prescriptivity', *Social Philosophy and Policy* 7, 151–174.

Railton, P. (1993): 'Reply to David Wiggins', in J. Haldane and C. Wright (eds.), *Reality, Representation, and Projection* (pp. 315–328), New York: Oxford University Press.

Shoemaker, S. (1981): 'Some Varieties of Functionalism', *Philosophical Topics* 12, 93–120.

Sturgeon, N. (1988): 'Moral Explanations', in G. Sayre-McCord (ed.), *Essays in Moral Realism* (pp. 229–255), Ithaca: Cornell University Press.

Department of Philosophy
New York University
100 Washington Square East
NY 10003-6688
USA
E-mail: dbb206@nyu.edu