Analytical Moral Functionalism Meets Moral Twin Earth

Terry Horgan, University of Arizona

Mark Timmons, University of Memphis

In Chapters 4 and 5 of his 1998 book *From Metaphysics to Ethics: A Defence of Conceptual Analysis*, Frank Jackson propounds and defends a form of moral realism that he calls both 'moral functionalism' and 'analytical descriptivism'. Here we argue that this metaethical position, which we will henceforth call 'analytical moral functionalism', is untenable. We do so by applying a generic thought-experimental deconstructive recipe that we have used before against other views that posit moral properties and identify them with certain natural properties, a recipe that we believe is applicable to virtually any metaphysically naturalist version of moral realism. The recipe deploys a scenario we call *Moral Twin Earth*.[1]

1. *Jackson's Analytical Moral Functionalism*[2]

We begin by briefly summarizing Jackson's theory. He proposes to construe moral terms like 'goodness' and 'rightness' in much the same way that the mentalistic terms of folk psychology are construed by the first-order version of the position in philosophy of mind called analytical functionalism. According to first-order analytical functionalism about the mental—as articulated, for instance, by D. M. Armstrong (1968, 1970) and David Lewis (1966, 1972, 1980)—mental-state terms are functionally definable via the principles of common-sense psychology, and these terms refer not to second-order functional properties but rather to certain neurophysical properties that fill the roles specified by the functional definitions.

Jackson's approach involves four central ideas. First, he posits what he calls 'folk morality', something whose (partly implicit) mastery he takes to be required for competence in the use of moral concepts:

> In the case of the mind, we have a network of interconnected and interdefinable concepts that get their identity through their place in the network…. The network itself is the theory known as folk psychology, a theory we have a partly tacit and partly explicit grasp of…. In the case of ethics, we have *folk morality*: the network of moral opinions, intuitions, principles, and concepts whose mastery is part and parcel of having a sense of what is right and wrong, and of being able to engage in meaningful debate about what ought to be done…. Moral functionalism, then, is the view that the meanings of the moral terms are given by their place in this network… (130)

Second, he claims that interpersonal commonality of meaning for moral terms requires that all parties are using moral terms in a way that reflects a mastery (partly implicit) of one and the same folk morality—which he calls *mature* folk morality. He also assumes that this prerequisite is satisfied in the case of humans, while acknowledging that if it is not then his account would need to be relativized:

> I have spoken as if there will be, at the end of the day, some sort of convergence of moral opinion in the sense that mature folk morality will be a single network…accepted by the community as a whole. Indeed, I take it that it is part of current folk morality that convergence will or would occur…. But this may turn out to be, as a matter of fact, false. Indeed, some hold that we know enough now about moral disagreement to know that convergence will (would) not occur. In this case, there will not be a single mature folk morality but rather different mature folk moralities for different groups in the community;

> and to the extent that they differ, the adherents of the different mature folk moralities will mean something different by the moral vocabulary…. I set this complication aside in what follows. I will assume what I hope and believe is the truth of the matter, namely that there will (would) be convergence. But if this is a mistake, what I say in what follows should be read has having implicit relativization clauses built into it. (137)

Third, he claims that moral terms have conceptual analyses that result by applying the method of defining theoretical terms developed by David Lewis (1970). The idea is first to characterize a system of properties that together conform to the principles of the theory—in this case, the principles of mature folk morality—and then to characterize each member of the system by its place within the whole. Jackson writes:

> Let **M** be mature folk morality. Imagine it written out as a long conjunction with the moral predicates written in property name style. For example, ‘Killing someone is typically wrong’ because ‘Killing typically has the property of being wrong’. Replace each distinct moral property term by a distinct variable to give $\mathbf{M}(x_1, x_2, \ldots)$. Then ‘$(\exists x_1) \ldots \mathbf{M}(x_1 \ldots)$’ is the Ramsey sentence of **M**, and
>
> $$(\exists x_1) \ldots (y_1) \ldots (\mathbf{M}(y_1 \ldots) \text{ iff } x_1 = y_1 \ \& \ x_2 = y_2 \ldots)$$
>
> is the modified Ramsey sentence of **M** which says that there is a unique realization of **M**.
>
> If moral functionalism is true, **M** and the modified Ramsey sentence of **M** say the same thing. For that is what holding that the ethical concepts are fixed by their place in the network comes to. Fairness is what fills the fairness role; rightness is what fills the rightness role; and so on. We can now say what it is for some action *A* to be, say, right, as follows:
>
> (R) A is right iff $(\exists x_1) \ldots (A \text{ has } x_r \ \& \ (y_1) \ldots (M(y_1, \ldots \text{ iff } x_1 = y_1 \ \& \ldots))$

> where 'x_r' replaced 'being right' in **M**. We now have our account of when *A* is right: it is right just in case it has the property that plays the rightness role as specified by the right hand side of (R)… (140-1)

Fourth, he maintains that certain first-order *natural* properties fill the respective roles that define the respective moral terms, and hence that moral terms *denote* these role-filling natural properties. (He remains neutral, and he says that mature folk morality is itself neutral, as to whether moral terms denote such role-filling natural properties rigidly or non-rigidly.) Here the key line of thought is this: by virtue of the a priori supervenience of the ethical on the descriptive, each moral term is necessarily coextensive with some natural property, viz., the disjunction of all natural supervenience-base properties for the moral term. That property is the role-filler, and hence is the referent-property of the moral term. (Jackson holds that necessarily-coextensive properties are identical, but he also offers independent arguments, including parsimony considerations, in support of the proposed property-identities; see pp. 125-8.)

2. Trouble: Moral Twin Earth[3]

Competent wielders of language and concepts have substantial intuitive mastery of the semantic norms governing the terms of the language they employ and the concepts those terms express—just as they have substantial intuitive mastery of the *syntactic* norms governing their language. Implicit mastery of the semantic workings of the term 'water' and the concept it expresses, for instance, presumably is reflected in people's strong intuitions about Putnam's Twin Earth scenario: e.g., the intuition that Twin Earthers do not mean by their

Twin English term 'water' what English speakers on Earth mean by 'water', and the intuition that the Twin English term 'water' is not translatable by the orthographically identical term of Earth English. Such intuitions constitute strong (though of course defeasible) empirical evidence for the hypothesis that ‘water’ *rigidly designates* the specific stuff called water here on Earth, viz., H20. For, this hypothesis nicely provides a plausible non-debunking explanation of the intuitions themselves—i.e., an explanation that treats the intuitions as veridical, and thus as reflective of people’s semantic competence with the concept WATER.

Presumably, competent wielders of *moral* terms and concepts have a comparable intuitive mastery of the semantic workings of ‘rightness’, ‘fairness’, and moral terms and concepts more generally. So if indeed these terms have conceptual analyses of the kind Jackson claims they do, 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) a plausible non-debunking explanation for these judgments is provided by analytical moral functionalism (henceforth, AMF).

Conversely, if the appropriate Twin Earth scenario does *not* possess feature (i)—i.e., if the semantic intuitions of competent speakers turn out not to be what they should be if AMF is true—then this will mean that AMF is probably false. We say ‘probably’ false because the inference to AMF’s falsity would be nondemonstrative, an inference to the best explanation. Semantic intuitions about Twin Earth scenarios are *empirical* evidence about matters of semantics (just as syntactic intuitions about grammaticality are empirical evidence about matters of syntax).[4]

We will now mount an argument against AMF by arguing that things go the latter way—i.e., that one’s intuitive judgments concerning a suitable Twin Earth scenario go contrary to AMF. For present purposes let us provisionally suppose that Jackson is right in his assumption that there is a single mature folk morality **M** to which all Earthly persons would converge under suitably ideal reflection. (We do not for a moment believe this supposition, and we will have more to say about it below. But applying our deconstructive recipe to any given version of naturalist moral realism involves granting any such optimistic assumption figuring in the view under consideration, and then arguing that the view would be mistaken even if the assumption were correct.) What is wanted, then, is a Twin Earth where things are as similar to earth as possible, consistent with the stipulation that there is some *different* mature folk morality **M***, distinct from **M**, to which all *Twin* Earthly persons would converge under suitably ideal reflection.

So let us begin by supposing that, as a matter of empirical fact, Earthly mature folk morality is consequentialist in nature, and is best systematized by some specific consequentialist normative theory; call this theory T^c. Let us further suppose that there is some reliable method of moral inquiry which, if properly and thoroughly employed, would lead Earth folks to discover this fact about their uses of moral terms and concepts.

Now consider Moral Twin Earth, which, as you might expect, is very much like good old Earth: same geography and natural surroundings, with people who live in Twin Australia and by and large speak Twin English, etc. Of particular importance here is that Moral Twin Earthers have a vocabulary that works very much like human moral vocabulary: they use the terms ‘good’ and ‘bad’, ‘right’ and ‘wrong’ to evaluate actions, persons, institutions, and so forth (at least those who speak Twin English use these terms, whereas those who speak some other Twin language use terms orthographically identical to the terms for good, etc., in the corresponding Earth language). In fact, were a group of explorers from Earth ever to visit Moral Twin Earth they would be strongly inclined to translate the Moral Twin Earth terms ‘good’, ‘right’, and the rest as identical to their own orthographically identical English terms. After all, the uses of these terms on Moral Twin Earth bear all the formal marks that are usually taken to characterize moral vocabulary and moral practice. In particular, the terms are used to reason about considerations bearing on the well being of persons on Moral Twin Earth; Moral Twin Earth people 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 twin-moral discourse and associated practice reveals that Twin Earthers all would converge, under ideal reflective inquiry, to a mature folk morality that is nonconsequentialist, and thus is distinct from the consequentialist mature folk morality to which (we are supposing) Earthers would all converge. Suppose too that Twin Earthly mature folk morality is best systematized by some specific deontological normative theory; call this T^d. The theory T^d, although importantly different from T^c, nonetheless is similar enough to 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 have already noted that if explorers from Earth ever visit Moral Twin Earth, they will be inclined, at least initially, to construe persons on Moral Twin Earth as having beliefs about good and right, and to translate Twin English uses of these terms into orthographically identical English terms.) The differences in the respective mature folk moralities of Earthers and Twin Earthers, we may suppose, are due at least in part to certain species-wide differences in psychological temperament that distinguish Earthers from Twin Earthers. (For instance, perhaps Twin Earthers tend to experience the sentiment of *guilt* more readily and more intensively, and tend to experience *sympathy* less readily and less intensively, than do Earthers.[5])

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’, ‘right’, ‘fair’, etc.? Two hermeneutic options are available. On one hand, one 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 Earthers designate the unique natural properties that respectively satisfy the respective Lewis-style conceptual analyses of those terms obtainable from theory T^c, whereas the twin-moral terms used by Twin Earthers designate *distinct* unique natural properties that respectively satisfy the respective conceptual analyses obtainable from T^d; hence, because corresponding moral and twin-moral terms have different, incompatible, conceptual analyses, moral and twin-moral terms *differ in meaning*, and are not intertranslatable. On the other hand, one 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 Earthers and Twin Earthers would be *genuine* disagreements—i.e., disagreements in moral belief and in normative moral theory, rather than differences in meaning.

We submit that by far the more natural and plausible 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 Earth uses of ‘good’ and ‘right’ as not translatable by the orthographically identical terms of English. But if AMF were true, and moral terms had the kinds of conceptual analyses that Jackson claims they do, 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 we have here a difference in meaning and that the twin-moral terms of Twin English are not translatable by English moral terms. But when it comes to characterizing the differences between Earthers and Moral Twin Earthers on this matter, by far the more 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 Earthers and Twin Earthers who have encountered each other. Suppose that Earthers visit Twin Earth (or vice versa), and both groups come to realize that members of their respective species would converge to different mature folk moralities that conform respectively to the consequentialist theory T^c and the deontological theory T^d. If AMF were true, then recognition of these differences ought to result in its seeming rather silly, to members of each group, to engage in inter-group debate about goodness—about whether it conforms to normative theory T^c or to T^d. (If, in Putnam's original scenario, the two groups learn that they have respectively been using 'water' to refer to two 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.

Since semantic norms are tapped by human linguistic and conceptual competence, and since the relevant competence is presumably reflected in one's intuitive judgments concerning Twin Earth scenarios, these intuitions about Moral Twin Earth constitute strong empirical evidence against AMF. Barring overwhelmingly strong reasons to think otherwise, the best explanation of the intuitive judgments is a non-debunking explanation that treats them as the products of semantic competence, and hence as veridical. And if indeed they are veridical, then analytical moral functionalism is false.

3. An Unappealing Fallback[6]

Briefly stated, the problem with Jackson's position is that it is guilty of *chauvinistic conceptual relativism*—that is, it is committed to the claiming that actual or possible agents who have a mature folk morality different from that of humans would not possess the concepts of goodness, rightness, etc. at all. This is objectionably human-centered, because it chauvinistically builds the folk morality supposedly shared by all of humankind directly into moral concepts themselves. In fact, Moral Twin Earthers would share moral concepts with Earthers, despite having somewhat different moral beliefs by virtue of their different mature folk morality. (By contrast, the claim that 'water' rigidly designates H20 is a *non-chauvinistic* form of conceptual relativism, because it does not wrongly exclude Putnam's non-human Twin Earthers from possessing a concept that they really do possess. On the contrary, the human concept WATER does indeed rigidly refer to the particular kind of clear liquid, whatever kind this is, that fills the lakes and streams in the local Earthly environment occupied by us humans. Likewise, the Twin Earthers' corresponding concept is indeed different, since it rigidly refers instead to the kind of clear liquid that fills the lakes and streams in the *their* local environment—viz., XYZ.)

The view in philosophy of mind often called *psychofunctionalism* is chauvinistic in a way similar to Jackson's metaethical position. According to psychofunctionalism, common-sense mentalistic concepts are functionally definable via the ideally complete and correct empirical psychological theory that is true of human beings—which we will call *mature empirical psychology*. But consider a race of possible creatures—say, Martians—who undergo internal states that (i) conform to all the principles of folk psychology just as much as the internal states of humans do, but (ii) conform to a somewhat different mature empirical psychology than do the internal states of humans. The trouble with psychofunctionalism is that it chauvinistically denies that such Martians have beliefs, desires, and other such mental states; it denies this despite the fact that Martians undergo internal states that jointly conform to the principles of folk psychology just as well as do the relevant internal states of humans themselves.

Analytical functionalism in philosophy of mind, on the other hand, avoids this form of conceptual chauvinism about mental states exhibited by psychofunctionalism. Analytical functionalism claims that the mentalistic concepts of folk psychology should be construed as being functionally definable not by the specific mature empirical psychology that happens to be true of humans, but rather by folk psychology itself. Analytical functionalism thus would count the lately-mentioned Martians as possessing beliefs, desires, and other folk-psychological mental states, by virtue of undergoing first-order states that collectively conform to the generalizations of folk psychology.[7] Analytical functionalism avoids the mistake of chauvinistically building into folk-psychological concepts the specific empirical psychological principles that happen to be true of humans.

In light of this comparative advantage of analytical functionalism vis-à-vis psychofunctionalism in the philosophy of mind, a fallback metaethical position that Jackson might consider would be to construe moral terms and concepts as functionally definable not by the specific mature folk morality supposedly possessed by all humankind, but rather by a set of uncontroversial *moral platitudes*: principles that are common currency within all possible mature folk moralities. This approach would avoid the problem of conceptual chauvinism.

But the trouble with this idea is that the kinds of platitudinous, non-tendentious, generalizations that clearly count as constitutive of people's common sense understanding of moral terms and concepts are simply not sufficient to pin down determinate referents for them. One can distinguish between *formal* and *substantive* moral platitudes. Formal moral platitudes would include those generalizations that link moral terms and concepts to one another and thus express definitional connections among them—for instance, "If an action is wrong, all things considered, then one ought not, all things considered, perform that action" and "If an action is morally permissible, all things considered, then it is not morally wrong, all things considered, to perform that action." There are also those formal moral platitudes that represent features of the so-called 'logic of moral discourse', like the principle of universalizability: "If an action is right (or wrong) for one agent to perform in certain circumstances, then it is right (or wrong) for any similar agent in similar circumstances." Substantive moral platitudes would be ones that apparently link moral concepts more directly to non-moral ones. Many philosophers have claimed that there are such substantive platitudes, for instance, "Right actions are concerned to promote or sustain or contribute in some way to human flourishing" and "Right actions are expressive of equal concern and respect."

But formal considerations alone clearly are not enough to secure determinate referents for moral terms and concepts; in general, such a priori constraints are compatible with any of a great variety of normative moral theories that deliver incompatible verdicts about numerous specific moral issues. Nor will appeal to substantive moral platitudes (together with the formal ones) suffice to produce referential determinacy. Consider, for example, the lately mentioned generalizations involving flourishing and impartiality. The generic notions of flourishing and impartiality are quite vague, and thereby can be construed very differently within competing, incompatible, moral theories. Let us focus for a moment on the notion of impartiality—the idea that everyone is to be accorded *equal respect*. The problem of appealing to this notion is that it lacks sufficient determinacy to serve as an anchor to uniquely pin down referents for moral terms. James Griffin brings out the point nicely:

> Every moral theory has the notion of equal respect at its heart: regarding each person as, on some sense, on an equal footing with every other one. Different moral theories parlay this vague notion into different conceptions. Ideas such as the Ideal Observer or the Ideal

> Contractor specify the notion a little further, but then they too are very vague and allow quite different moral theories to be got out of them. And the moral theories are not simply derivations from these vague notions, because the notions are too vague to allow anything as tight as a derivation. Too vague, but not totally empty; although the moral theories that we end up with put content into all these notions, the notions themselves also do something toward shaping the theories. (Griffin 1986, 208)

Talk of flourishing is vague in just the same way. Moreover, the same will be true of other notions that, like equal respect and flourishing, might plausibly be understood as part of the very concept of moral thought and discourse. So it does not appear that moral platitudes alone can collectively generate Lewis-style functional definitions that fix determinate referents for terms like 'goodness', 'rightness', and 'fairness'. As Michael Smith (1994) remarks, "These platitudes need not and should not be thought of fixing a unique content or substance for moral reasons all by themselves, rather they simply serve to tell us when we are in the ballpark of moral reasons, as opposed to the ballpark of non-moral reasons" (p. 184); for further substantiation of this claim, see Smith's own discussion.

One might try maintaining (i) that the indeterminacy here described involves only relatively few borderline hard cases about which competing moral theories would disagree, and (ii) that these cases can be comfortably relegated to the category "no moral fact of the matter." But actually the resulting indeterminacy of truth-value would be massive, since it would extend to virtually any kind of case about which there is actual or potential moral disagreement. Acceptance of similar-looking, superficially substantive, platitudinous principles by people with differing moral values does very little to secure agreement about concrete cases, because the concepts that feature in such principles—equal respect, for instance—are apt to be applied to specific acts and situations so very differently by the different parties. (For a powerful elaboration and defense of this claim, see Snare 1980.)

Thus, the fallback retreat that replaces the appeal to mature folk morality by an appeal to uncontroversial moral platitudes is not viable, because it immediately encounters—with a vengeance—the problem of radical indeterminacy of reference for moral terms and radical indeterminacy of truth-value for moral statements. Out of the frying pan of chauvinistic conceptual relativism, into the fire of radical moral indeterminacy! This is an instance of a generic dilemma posed by Moral Twin Earth for naturalist versions of moral realism: objectionable relativism on one hand, or objectionable indeterminacy on the other.

4. The Full Extent of Jackson's Chauvinistic Conceptual Relativism

Earlier we granted, for the sake of argument, Jackson's optimistic assumption that there is a single mature folk morality to which all humans would converge under ideal reflection. In adapting our generic recipe for cooking up specific Moral Twin Earth counterexamples against specific versions of naturalist moral realism, we argued that even if the optimistic assumption is true, Jackson's analytic moral functionalism is untenable anyway; for, it chauvinistically builds into moral concepts the specific mature folk morality that supposedly would be converged upon by all *Earthers*, thereby wrongly entailing that Moral Twin Earthers lack the moral concepts that we Earthers possess.

But it is entirely possible—we think likely—that different *humans* would converge to different mature folk moralities under ideal reflection. Prima facie, this is the most plausible explanation of the persistent, recalcitrant-looking, actual moral disagreements that commonly exist within humankind. As we pointed out in section 1, Jackson is unflinching in the face of this possibility: he is prepared to extend his conceptual relativism to different human subgroups, if necessary. As he says (in a passage quoted earlier) about the possibility that convergence to a single folk morality would not occur, "In this case, there will be not a single mature folk morality but rather different mature folk moralities for different groups in the community; and to the extent that they differ, the adherents of the different mature folk moralities will mean something different by the moral vocabulary" (137).

This is a very large bullet to bite. Jackson's position entails that apparent moral disagreements among humans with deeply differing moral values are *merely* apparent: the different parties are expressing different concepts with their moral terms, are talking past one another rather than disagreeing, and often are both right given what they respectively mean by their moral terms. This conceptual-relativist construal of such apparent moral disagreements is *wildly* contrary to the common sense, intuitive, way of understanding such situations. Common sense, and ordinary discursive practice, construe the appearances as veridical: the parties in such a dispute are employing common moral concepts, are using moral terms with common meaning, and are engaged in a deep and genuine disagreement in *moral belief*. Barring some overwhelmingly strong reason to think that this common sense construal of such cases is mistaken, the enormous size of the bullet Jackson is biting constitutes a further strong consideration against his position. (This is an extension of, and a further strengthening of, the lesson of Moral Twin Earth; the point is that deep moral disagreements of the kind described in the Moral Twin Earth scenario very probably exist right here on Earth.)

5. Non-Descriptivist Cognitivism vs. Analytical Moral Functionalism

We do not claim to have conclusively refuted Jackson's metaethical position; conclusive arguments are rare in philosophy. Philosophical theories, like scientific theories, should be evaluated in terms of their overall theoretical benefits and costs—and so should be evaluated *comparatively*, with an eye on benefits and costs of the competing philosophical theories on the conceptual landscape. Bullet-biting can be appropriate, if the advocates of competing theories all must bite even bigger bullets. Thus, how telling our negative arguments are against Jackson ultimately depends in part upon what available alternative metaethical theories exist, and upon the viability of those alternatives. In this section we locate our case against AMF within a wider dialectical setting, by briefly comparing it with the metaethical position we ourselves favor—a version of non-descriptivism.

Non-descriptivists maintain that the overall declarative content of a moral judgment or a moral sentence is not descriptive content: such a judgment or sentence does not represent the world as being a certain way. Jackson, following the usual tradition in metaethics, uses the term 'non-cognitivism' for non-descriptivism. He says this about non-descriptivism, thus labeled:

> It is only under the assumption of cognitivism that ethics presents a location problem. If the non-cognitivists are right and ethical sentences do not represent things as being a certain way, there is no question of how to locate the way they represent things as being in relation to how accounts told in other terms—descriptive, physical, social, or whatever—represent things as being…. Although I cannot rule out non-cognitivism simply by noting that ethical sentences are meaningful and syntactically right for truth, I do think it is very much a 'last resort' position. (117)

Elsewhere (Timmons 1999 chapter 4, Horgan and Timmons 2000b, in press a, in press b, forthcoming), we ourselves have articulated and defended a version of non-descriptivism that we maintain is not a 'last resort' position at all, but rather is more plausible than the various other metathical positions currently on offer. We call this view both 'non-descriptivist cognitivism' and ‘cognitivist expressivism’; here we adopt the former name. We now present an extremely truncated sketch of this position.

Non-descriptivist cognitivism makes the following claims. (1) Contrary to non-cognitivist views like emotivism and prescriptivism, moral judgments are genuine *beliefs*, and utterances of moral sentences are genuine *assertions*. (The label ‘non-cognitivism’ fits emotivism and prescriptivism because these views deny that moral judgments are beliefs, and instead treat them as non-cognitive states—for instance, as *conative* states of approval or disapproval.) (2) Contrary to descriptivist views, the overall declarative content of moral beliefs and assertions is not descriptive content: these beliefs and assertions do not represent things as being a certain way. (3) Beliefs with the most basic kinds of declarative content are *psychological commitment-states* with respect to a logically atomic descriptive content. (4) Such commitment-states are of two fundamental kinds: *is*-commitments and *ought*-commitments; each kind has both affirmative and negative versions. Thus, there are four basic kinds of belief: *affirmative is-commitments* (e.g., the belief that Bush is U.S. President), *affirmative ought-commitments* (e.g., the belief that it ought to be that Gore is U.S. President), *negative is-commitments* (e.g., the belief that it’s not the case that Gore is U.S. President), and *negative ought-commitments* (e.g., the belief that it ought not to be the case that Bush is U.S. President). (5) The constitutive features of ought-commitments include both (a) certain distinctive typical roles played by these states in the psychological economy of the morally-judging agent, including distinctively motivational roles in the case of first-person ought-commitments, and (b) certain distinctive typical phenomenological features exhibited by these states, including an experiential aspect of “reasons-based fittingness.” (6) The constitutive features of logically complex beliefs with component moral content—e.g., the belief that *if* Jones promised his wife to pick up the children from school *then* Jones ought to pick up the children from school—involve certain distinctive *inferential roles* played by such states in an agent's psychological economy, roles in which various kinds of logically basic is-commitments and/or ought-commitments are "in the offing." (7) A sincere assertion that p *expresses* the belief that p but does not *describe* that belief; thus, when the belief expressed is an ought-commitment, the declarative content of the corresponding assertion is not descriptive content. (8) Moral assertions typically have action-guiding roles in social intercourse that are similar to the typical action-guiding roles of moral beliefs (especially first-person moral beliefs) in a morally judging agent's own psychological economy. (9) A truth ascription to a moral belief or statement normally conforms to the Tarski T-schema, and thus normally constitutes a fusion of moral and semantic evaluation; such a morally engaged truth ascription is a meta-level expression of a moral belief (i.e., an ought-commitment), and hence is itself non-descriptive in overall declarative content (as is the first-order moral judgment or statement to which truth is ascribed).[8]

If indeed non-descriptivist cognitivism is a viable and independently plausible metaethical position, as we argue in the papers lately cited, then this fact considerably strengthens the force of Moral Twin Earth scenarios as evidence against naturalist moral realism in its various versions, including Jackson's version. Non-descriptivism is not a 'last resort' position that is worth avoiding even at the cost of embracing chauvinistic conceptual relativism.

One final point. The availability of non-descriptivist cognitivism as a credible theoretical option also calls into question a dialectical move that Jackson makes in an effort to fend off a classic argument against naturalist moral realism, G. E. Moore's famous "open question argument." Jackson says:

> It may be objected that even when all the negotiation and critical reflection is over and we have arrived at mature folk morality, it would still make perfect sense to doubt that

> the right is what occupies the rightness role. But now I think that we analytical descriptivists are entitled to dig in our heels and insist that the idea that what fits the bill *that* well might still fail to be rightness, is nothing more than a hangover from the platonist conception that the meaning of the term 'right' is somehow a matter of its picking out, or being somehow mysteriously attached to, the form of the right. (151)

But of course the open question argument has long been employed by *non-descriptivists*, and not merely by non-naturalist moral realists, against naturalist moral realism. The argument has considerable intuitive force, and indeed is closely related in spirit to our own Moral Twin Earth argument; see Horgan and Timmons (1992a). Even granting the (dubious) assumption that all humans would converge upon a single mature folk morality, there is nothing especially platonistic about the claim that some possible agent who employs the same moral concepts that humans do—e.g., a Moral Twin Earther—could intelligibly doubt, of a natural property that the agent knows fits the rightness role that is functionally defined by the mature folk morality *of humans*, whether this natural property is identical to rightness. On the contrary, to insist that there could be no such moral agent is to be guilty of conceptual chauvinism.

[1] See Horgan and Timmons (1991, 1992a, 1992b, 1996a, 1996b, 2000a) and Timmons (1999 chapter 2).

[2] All citations to Jackson, in this section and throughout the paper, refer to Jackson (1998).

[3] This section is largely adapted from section IV of Horgan and Timmons (1992a), with minor modifications to make the discussion directly applicable to Jackson's position.

[4] For more on philosophical appeals to intuition as being, in effect, semantic-competence arguments that provide empirical support for philosophical hypotheses about the semantics of concepts and terms (even though philosophers often do not appreciate this fact about their own intuition-based reasoning), see Horgan (1993) and Graham and Horgan (1994). For treatment of such arguments as conforming to part, but not all, of the traditional conception of the a priori, see Henderson and Horgan (2000, 2001), where

these arguments are dubbed "low-grade a priori" because they rest on data that is armchair-accessible, such as one's own semantic intuitions.

[5] In order to forestall any attempt to parlay this postulated difference into a basis for resisting the argument we are about to give, let us further stipulate (i) that the difference is merely a matter of initial psychological *tendencies* within Twin Earthers and Earthers respectively, (ii) that these tendencies are psychologically malleable in both groups, and thus (iii) that both groups are *plastic* with respect to how their moral sensibilities get molded, rather than being "hard-wired." Thus, for both groups it is true that if certain cultural developments *were* to transpire, then the members of the group *would* develop an altered moral sensibility and would sustain this change via alterations in moral education. For instance, if someone like Peter Singer were to exert widespread influence on Moral Twin Earth, then the Moral Twin Earthers would develop and sustain a utilitarian moral sensibility. Or, if the concept of sin were to become ubiquitously influential on Earth, then Earthers would develop and sustain a deontological moral sensibility. (We ourselves would argue not only that human moral psychology is indeed malleable in this way, but also that such differences in moral sensibilities and in associated modes of moral education are abundantly present right here on Earth. But remember that we are currently granting, for the sake of argument, Jackson's optimistic assumption that there is a single mature folk morality to which all Earthers would converge under ideal reflection.)

[6] Parts of this section are adapted, with minor modifications, from section 5 of Horgan and Timmons (1996a) and from section 2 of Horgan and Timmons (2000a).

[7] Assuming that the relevant first-order states are different in Martians than in humans (say, because Martians are composed of silicon rather than organic molecules), a first-order version of analytical functionalism will need to construe mental state-names as *population-specific nonrigid designators* in order to accommodate Martian mentality. David Lewis explicitly took this tack in Lewis (1980).

[8] Non-descriptivist cognitivism does not claim, however, that this morally engaged way of using the truth predicate is the only legitimate use. We ourselves maintain that the concept of truth is governed by

implicit, contextually variable, semantic parameters, and that in the case of moral beliefs and assertions, any of three distinct uses of the truth predicate can be semantically sanctioned in a specific context: (1) a morally engaged *disquotational* use, expressive of one's own moral beliefs, (2) a morally disengaged, nondisquotational, *correspondence* use, under which only beliefs and assertions whose overall declarative content is descriptive can be true or false, and (3) a morally disengaged, nondisquotational, *overtly relativistic* use, under which truth ascriptions get explicitly relativized to the moral standards of some specific person or group. When non-descriptivists assert that moral judgments and statements are neither true nor false (as we ourselves sometimes do), the truth predicate is being employed in manner (2) rather than manner (1). This is not inconsistent with using the truth predicate disquotationally vis-à-vis moral claims, although one cannot use it both ways in one breath. Also, when the truch predicate is used in the third, overtly relativized, manner, typically one is simply making a descriptive remark about what the standards of some person or group imply about the moral status of a type of action or an act token. Again, using the truth predicate this way on one occasion is not inconsistent with using it disquotationally (or correspondence-wise) on another. For further related discussion, see Timmons (1999 chapter 4), especially pp. 149-52, Horgan (2001 section 5), Horgan and Timmons (2000b section VII.1, 2002 notes 10 and 19, and forthcoming).