Problems for Moral Twin Earth Arguments
Author(s): Joshua Gert
Source: *Synthese*, Vol. 150, No. 2 (May, 2006), pp. 171-183
Published by: Springer
Stable URL: http://www.jstor.org/stable/20118766
Accessed: 16/02/2014 20:17

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
http://www.jstor.org/page/info/about/policies/terms.jsp

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Springer is collaborating with JSTOR to digitize, preserve and extend access to *Synthese.*

http://www.jstor.org

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

JOSHUA GERT

PROBLEMS FOR MORAL TWIN EARTH ARGUMENTS[☆]

ABSTRACT. Terry Horgan and Mark Timmons have recently presented a series of papers in which they argue against what has come to be called the 'new wave' moral realism and moral semantics of David Brink, Richard Boyd, Peter Railton, and a number of other philosophers. The central idea behind Horgan and Timmons's criticism of these 'new wave' theories has been extended by Sean Holland to include the sort of realism that drops out of response-dependent accounts that make use of an analogy between moral properties and secondary qualities. This paper argues that Holland's extension depends crucially on the fact that his target is a *direct* response-dependent account of *moral* value. His argument does not work against such accounts of more basic normative notions such as 'harm' or 'benefit'. And these more basic notions may then serve as the basic normative building blocks for an indirectly response-dependent moral theory.

Terry Horgan and Mark Timmons have recently presented a series of papers in which they argue against what has come to be called the 'new wave' moral realism and moral semantics of David Brink, Richard Boyd, Peter Railton, and a number of other philosophers.[1] The moral realism at issue for Horgan and Timmons does not depend in any way on a purported analogy between moral properties and secondary qualities. Rather, it depends on a purported causal role for moral qualities in explaining our use of moral terms – a role similar to that played by H_2O in explaining our use of the term 'water'. However, the central idea behind Horgan and Timmons's criticism of these 'new wave' theories has been extended by Sean Holland to include the sort of realism that emerges from response-dependent accounts that do make use of the analogy between moral properties and secondary qualities.[2] This paper, as a contribution to the program of constructing a viable normative realism on a response-dependent basis, argues that Horgan and Timmons's argument, and Holland's extension of it, depend crucially on the fact that the target is a response-dependent account of *moral* value. Their arguments do not work against response-dependent accounts of more basic normative notions such as 'harm' or 'benefit'. And since it may well be possible to use the notions of harm and benefit as the basic nor-

Synthese (2006) 150: 171–183
DOI 10.1007/s11229-004-6262-y
© Springer 2006

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

mative building blocks for a conceptually complex moral theory, there is a sense in which their arguments do not undermine response-dependent accounts of morality either.[3]

Horgan and Timmons's argument against new wave moral realism is briefly as follows. The new wave moral realist holds that the reference fixing by which 'water' comes to refer – rigidly – to H_2O is of the same sort by which 'morally good' comes to refer – also rigidly – to whatever naturalistic property it might be that serves to explain our use of that phrase. That is, even before we humans knew that water was made up of H_2O, it was nevertheless true that H_2O was playing a certain sort of causal role in the processes by which we managed to develop a word to refer to the clear wet stuff that fills lakes and streams, etc. Because H_2O was playing this role, it was the referent of the word 'water', even before we knew anything about chemistry.[4] When we discovered that water was made of H_2O, we learned a necessary *a posteriori* truth. Because the proposition 'water is H_2O' is a *necessary* truth, the word 'water' (as used by us) refers to H_2O even in possible worlds where the clear wet stuff that fills lakes and streams is made up of XYZ instead of H_2O. In one of those worlds – call it 'Twin Earth' – there may be people who speak something very much like English, and who have a word, 'water' that refers to the clear wet stuff in *their* lakes and streams. But because the causal explanation of their use of the word 'water' gives a certain causal role to XYZ and not to H_2O, it follows that their word 'water' refers to XYZ and not H_2O.[5]

Horgan and Timmons argue against the idea that the above sort of account of the semantics of 'water' can be applied with equal plausibility to moral terms. Why is this? When visitors from Earth voyage to Twin Earth, bringing bottles of water, both Earthlings and Twin Earthlings may initially think that they are all using the word 'water' with the same meaning. In the laboratory, however, they may discover that the stuff in the Earthling's bottles is H_2O, whereas the stuff that comes out of the taps is XYZ. Upon making this discovery, they will (it is argued) see that Earthlings use the word 'water' with a different referent and therefore a different meaning than do Twin Earthlings. There will then be no further argument about what 'water' "really" means. But now consider what would happen if we tried to give the same sort of account of the semantics of 'morally good'. Suppose that at some distant happy point on Earth, moral philosophy has come to an end, and it has been discovered that the naturalistic property that plays the right sort of causal role in explaining our use of the phrase 'morally

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

good' is the property of maximizing overall happiness. And suppose that on Twin Earth, their philosophers have made an equally clear and irrefutable case for the claim that the naturalistic property that plays the right sort of causal role in explaining *their* use of the phrase 'morally good' is the property of having been done from the motive of duty. Suppose further that during a visit by Earthlings to Twin Earth, one of the visitors performs an action that clearly maximizes overall happiness, but that, equally clearly, could not possibly have been performed from the motive of duty. It does not seem plausible to Horgan and Timmons that the kind of dispute to which this action might be expected to give rise would be dissolved, as it was in the case of water, by the realization that Earthlings and Twin Earthlings use the phrase 'morally good' to refer to distinct naturalistic properties. Rather, Horgan and Timmons expect that the argument that will ensue will look just like the arguments that currently happen here on Earth between consequentialists and deontologists: all parties to the dispute will think that they are talking about the *same* property, and that other parties have *false beliefs* about the property. If so, there seems to be a significant problem in telling the same sort of semantic story about 'morally good' as we told about 'water'.

The root of the above problem for new wave moral realists is that their account ties the meaning of a moral term to its referent so closely that there cannot be a change in referent without a change in meaning.[6] This is the basic point from which Horgan and Timmons go on to infer that, according to new wave moral realism, we must regard Earthlings and Twin Earthlings as talking past one another whenever they seem to be having a moral dispute. For if Earthlings and Twin Earthlings use the phrase 'morally good' to refer to different naturalistic properties, then the new wave moral realist seems committed to the claim that they are also using that phrase with two distinct meanings. Hence, their apparent disputes about moral goodness are *only* apparent.

Holland extends the scope of Horgan and Timmons's argument by pointing out that response-dependent accounts, when they are corrected to avoid a certain objection, end up tying the meaning of moral terms to their referents just as tightly as do new wave moral realists. And as a result they become equally vulnerable to Twin Earth counterexamples.[7] How does this happen? Consider a response-dependent account of redness, along the lines of the following:

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

(1) x is red $\leftrightarrow$ normal subjects would, under perceptually normal conditions, experience x as red.

Now, 'normal subjects' cannot here be defined as 'subjects constituted so that, under perceptually normal conditions, they experience as red all and only red objects'. That would trivialize (1). But, as Crispin Wright believes (and as is plausible), 'normal subjects' might, in this context, be given a statistical reading.[8] After all, it is plausible that an object is red if and only if a very high percentage of human beings would, under perceptually normal conditions, experience it as red.[9] However, if we do give 'normal subjects' this statistical reading, then (1) stands exposed to the following objection: the statistics may change, but even if they did, the colors of object would not. That is, suppose that the end of human life is fast approaching, and that only one human subject remains alive to note the colors of objects. If this last human should turn out to be color-blind, this will not affect the colors of stones on the far side of Mars (nor, for that matter, the colors of the stones that this last human misperceives). How can a response-dependent account avoid this problem? One now-standard way is to rigidify the referent of 'red' by inserting an 'actually' operator into the analysis. Thus, (1) might become the following:

(2) x is red $\leftrightarrow$ the overwhelming majority of subjects, as they actually are in this world now, would, under perceptually normal conditions, experience x as red.

(2) fixes the referent of 'red' in a way that assures us that no subsequent variation in the visual apparatus of human beings can be relevant to the colors of objects. But with the introduction of a reference to the way in which most human being *actually* are, the referent of a response-dependent term becomes an unchanging naturalistic property, as it was for the new wave moral realist. If a response-dependent account of 'morally good' yields different referents on Earth and on Twin Earth, then it seems that such an account is vulnerable to the same objection that Horgan and Timmons level at new wave moral realists. The advocate of a response-dependent account will be forced to conclude that Earthlings and Twin Earthlings are only talking past one another when they seem to be engaged in a dispute of the sort that, here on Earth, we would recognize as the perennial, and genuine, dispute between consequentialists and deontologists.

It is the central point of this paper to suggest that the advocate of a response-dependent account of value can avoid the force of the above

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

objection by shifting the target of analysis from complex normative concepts such as 'moral goodness' or 'moral wrongness' to more basic concepts such as 'harm' and 'benefit'. No one would think it plausible to offer a response-dependent account of 'covered in purple and yellow blobs', although this is a visual category. Rather, someone who wished to offer a response-dependent account of visual properties would offer such an account only of basic visual concepts, and then use those concepts, along with others, in order to explain what it is to be, for example, covered in purple and yellow blobs. This is what philosophers who are drawn to the analogy between value and secondary properties should also do. They should provide response-dependent accounts of the basic normative concepts, and then use those concepts, together with other (non-normative) ones, to define more complex normative concepts such as moral rightness and wrongness.

In order to make my point more plausible, it may first be useful to draw a distinction between what we might call 'harm, broadly speaking', and 'basic harm'. It is relatively uncontroversial that when one is hit by a car, one suffers a harm. So too does one suffer a harm when one is put in jail. The limitless possibilities for extending this list might initially suggest that there are an infinite variety of harms. But in each of these cases, and in many others, it makes sense to ask 'But why is it harmful?' There are answers to this question: 'because it hurts', 'because it increases one's likelihood of an early death', 'because it increases one's likelihood of losing some abilities', and so on. However, if one asks the same sort of question about each of these more basic harms, eventually one comes to a point where all one can say is something like 'What do you mean, "Why is pain a harm?"' When one has come to this point, one has identified a *basic harm*, and there is no explaining, in more basic normative terms, why something counts as a basic harm.[10] In parallel fashion, for the notion of aversion, there are what one might call 'motivated aversions' and 'basic aversions'. Roughly put, one has a motivated aversion to an item when there is an answer to the question 'Why are you averse to that item?', while one has a basic aversion if the most appropriate response to such an question is something like 'What do you mean, "Why am I averse?"' A response-dependent account of 'basic harm' might go something like the following.

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

(3) x is a basic harm $\leftrightarrow$ the overwhelming majority of subjects, as they actually are in this world now, have a basic aversion to x.

Various other senses of 'harm' and its cognates can then be defined in terms of this notion, and will inherit its objectivity. For example, we can define 'harmful' as 'increasing the likelihood of suffering a basic harm', and 'harm' as 'either a basic harm, or something that causes or encompasses a basic harm'. Thus we could say that losing one's job usually counts as a harm.[11]

In fact, a slightly more sophisticated response-dependent account of basic harm would not take (3) as a *definition* of 'basic harm'. Rather, it would take (3) as an empirical truth that is useful in explaining why the concept 'basic harm' has, as its referent, those consequences of action that happen to produce basic aversion in the overwhelming majority of actual subjects. That is, such an account would start by drawing attention to the fact that there are consequences of action that, as a matter of fact, produce basic aversion in the overwhelming majority of subjects. This overwhelming agreement in attitude is what allows for the ostensive teaching of the concept 'harm' in such a way that it includes a basic core that has an objective naturalistic referent. For the existence of overwhelming agreement means that virtually all language speakers will go on in the same way after the requisite amount of ostensive teaching and correction of uses of the word 'harm'. And it means that those who do *not* go on in the same way will be corrected, so that even they will learn that 'harm' has a referent, and cannot be used simply to label the things to which *they* happen to be averse. In this way ostensive teaching parlays "mere" overwhelming agreement in phenomenological or affective response into a more complete agreement in meaning – even among people who do not share the characteristic response. For even those with affective disorders are constrained to use the word 'harm' with the univocal meaning it has in the language. This is also why color words have objective referents despite the existence of color-blind people.

The appeal to the notion of a *basic* aversion in the above account of 'basic harm' helps the account avoid the following sort of objection. Suppose that in some poor and drought-plagued land the custom has developed of using a mixture of cow blood and cow urine as a food source. Let us also suppose that the vast majority of actual humans would be averse to eating this mixture. Despite this nearly universal aversion, it sounds rather imperialistic to describe what these minority people are doing as 'willingly harming themselves'.

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

This is because they can provide answers to the question 'Why are you eating that?' Here are some plausible answers: 'Because otherwise I will become weak from hunger', 'Because otherwise I risk illness from malnutrition', 'Because it is my principle food source, and I will probably die if I don't eat it'. And follow-up questions such as 'Why are you averse to illness, physical weakness, and death?' (and any subsequent follow-up questions prompted by the answers) will eventually bottom out by revealing basic aversions, such as aversions to death, pain and disability. Since aversions to these kinds of things are almost universal, it is plausible to regard them as basic harms. Moreover, when one asks a typical citizen of the US why he would be averse to eating a mixture of cow blood and cow urine, there will be answers of the following sort: 'Because it will make me physically sick', 'Because it is going to taste terrible'. Let us suppose that these answers are actually *correct* for the US citizen. It is plausible that these aversions are basic, and that they correspond to basic harms. So it will turn out that eating this mixture will be harmful for most readers of this article, but will not be harmful for the members of the tribe who are used to eating it. There is nothing especially troubling or relativistic in this conclusion. In a similar way, for someone who is allergic to peanuts, eating them will cause harms. For a normal person, eating them will not. Similar remarks serve to deflect a related objection: that there may be things to which the majority of people are *not* averse, but which we all admit are harmful: Twinkies, for example. If (as is plausible) Twinkies are in fact harmful, it is surely because they lead to consequences such as the following: indigestion, heart disease, increased risk of death, the inconveniences attendant on obesity, and so on.[12] And the vast majority of people *do* have basic aversion to these things. So the suggested account will allow us to say that eating Twinkies causes harms (colloquially: 'that Twinkies are harmful') despite the failure of many people to be averse to them, on the whole.

Admittedly, the above account of harm is frankly foundational, and will therefore not appeal to all philosophical tastes. But such a response-dependent account of harm offers an *explanation* for foundationalism about harm. Such an account explains why basic harms can be taken to include all and only those things to which the overwhelming majority of human beings have a basic aversion, and why harms more generally can be understood to "bottom out" in basic harms. One cannot offer *reasons* why something should count as a basic harm, in the way that one can offer reasons why some action

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

should count as morally good.[13] Of course one can offer a semantic theory that *explains why* the term 'harm' includes, in its extension, a certain substantive consequence – that is, after all, the point of giving a response-dependent account, or some other genealogical account of the development of the term 'harm'. And one might then misleadingly go on to make the following claim: pain is harmful because almost everyone is averse to it. This sounds rather similar to the claim that being burnt is harmful because it causes a lot of pain. But the former claim concerns the determination of the sense of 'harm', while the latter *employs* that sense.[14]

Because we cannot offer reasons why any given basic harm should count as a basic harm, any more that we can offer reasons why something that looks *like that* should count as looking red, we will never see the kind of articulate, intelligible dispute about basic harms that we often see between consequentialists and deontologists, or in moral argument between Earthlings and Twin Earthlings. Rather, the dispute about basic harm will simply be between those who assert 'Pain is a harm' and those who deny it. This sort of irresolvable dispute *does* have the flavor of one in which the disputants are simply talking past one another. Of course, one or both of the disputants may think that there is a unique correct answer to the question, given by God or by facts about Platonic heaven. But that does very little to suggest that they are actually talking about the same thing.[15] In the case of 'harm', the dispute may become more heated than in the case of 'water' or 'red'. But given that the person who sincerely utters 'x is a harm' has a strong affective response to the prospect of getting x for himself or those he cares about, this is not surprising. Even if one does not accept C.L. Stevenson's account of emotive meaning, one should certainly accept his point that there can be disagreements in attitude that do not involve disagreement in belief.[16] And in this case, Stevenson may offer the best interpretation of the disagreement between those who can sincerely utter 'x is a harm', and those who cannot.

It is true that the argument offered in this paper suggests that any significant and consistent disagreement in sincere applications of basic terms such as 'harm' and 'benefit' would imply a conceptual disparity. But suppose for the moment that there are no such disagreements, and that there is no such disparity between the Earthling and Twin Earthling concepts of *harm* and *benefit*: that Twin Earthlings use the words 'harm' and 'benefit' with the same meanings that we do. And suppose that is because the response-dependent account

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

of harm and benefit suggested in this paper is correct, and Earthlings and Twin Earthlings share basic aversions. Finally, suppose that despite this agreement in the meanings of 'harm' and 'benefit', all Earthlings have become convinced that a particular consequentialist moral theory is correct, while all Twin Earthlings hold to a particular deontological theory. Perhaps this is the result of some subtle differences in our respective cognitive setups. Need the advocate of a response-dependent account of 'harm' say that our apparent moral disagreements are *only* apparent, and that despite our agreement about what 'harm' means, we disagree about what 'morally good' means? No, he need not. The concepts of 'morally good' and 'morally bad' are sufficiently complex that it is not surprising when people come up with different theories about them. When one comes up with such a theory, one is trying to explain the nature of moral goodness and badness in a way that is (as far as possible) consistent with our moral intuitions, and that also gives satisfying answers to questions such as 'What are we talking about, when we make moral judgments?' and 'What reasons are there for doing morally good things, and for avoiding morally bad ones?' It is certainly possible for two people engaged in this very project to come up with different theories. When this happens, and when their theories yield different moral verdicts about particular actions – verdicts that they take to be correct – it is appropriate to say that they have a substantive disagreement about the moral status of those actions. It makes sense to say this about two moral philosophers on Earth, and it makes sense to say this about an Earthling and a Twin Earthling. Since a response-dependent account of basic normative terms allows for this possibility, it provides a way to demystify the domain of the normative – a way to explain the basic normative building blocks in terms of which all other normative notions are to be analyzed – that is invulnerable to moral Twin Earth arguments.

Although it is of course well beyond the scope of this paper to show precisely how one would provide a full account of morality using a small handful of normative primitives, such as harm and benefit, it *is* part of the point of this paper to provide reasons for optimism about the prospects of providing one. I will do so by providing a sketch of one way that such an account might be constructed. Suppose, then, that we have a viable response-dependent account of 'harm' and 'benefit'. How might we develop this into an account of other normative notions, and ultimately into an account of morality? We could start by giving an account of practical reasons

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

in terms of harms or benefits: one has a reason to perform A, rather than B, in virtue of the fact that doing so will avoid or diminish the risk of a harm, or produce or increase the likelihood of a benefit, for some person. The adequacy of such an account would depend upon, amongst other things, its squaring with all the kinds of things we want to say about reasons for action, including the substantive things we want to say, as that one always has at least some reason to get out of a burning building. But one kind of criticism that such an account will not need to fend off is that it ignores or fails to account for the normative aspect of reasons for action. For the account is explicitly offered in terms of the normative notions of harm and benefit. And it is typically this charge that produces the biggest troubles for philosophers. Note that on this definition of 'reason for action', an agent has a reason against a certain action even if the foreseeable harm his action will produce is not a harm that *he himself* will suffer. Rather, on this account of reasons for action, an agent has a reason to avoid harming other people, even if the agent does not particularly care about those people.[17]

Now, with such an account of practical reasons, we might be able to go on to give an account of practical rationality. And with an account of practical rationality, we might be able to go on and give an account of morality. At each step, of course, more conceptual apparatus will have to be added, and the concepts will become increasingly complex. But this increasing complexity does not imply that any new normative primitives are being used. For example, the move from a definition of practical rationality to morality might be something like the following. An action is immoral if it breaks one of the rules that impartial rational people would want to have as part of an enforceable system, when there are no special circumstances such that impartial rational people would want it to be known that the rule could be broken in those circumstances.[18]

It is true that the notion of impartiality has an aura of normativity about it, so one might worry that new normative notions are being introduced at this point. But the impression that impartiality is a normative notion may be a result of the fact that many moral duties require impartiality, so that a failure to be impartial is often a moral failure.[19] But sometimes impartiality, rather than partiality, is morally suspect; this is often the case when people have legitimate expectations that we will be partial towards them. And it is possible for a sadistic tyrant to be impartial in the way in which he distributes undeserved suffering. So it is possible that impartiality could be

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

described in non-normative terms. But even if impartiality is in fact a normative notion, this may be only because it involves the notions of 'harm' and 'benefit'. For example, one plausible account of impartiality holds that to be partial is to base one's decisions or actions partly on consideration of who will be harmed or benefited by them.[20] This chain of definitions shows how a response-dependent approach to the normative might underwrite a completely naturalistic account of morality.

It is undeniable that the approach advocated in this paper favors theories that have a certain consequentialist element. This is true despite the fact that the actual moral rules that emerge need be no more consequentialist than those that belong to typical deontological views. However, it is unclear if this consequentialist bias can form the basis of a criticism, even if the consequentialism emerged more starkly in each of the final moral rules. For if it turns out that 'harm' and 'benefit' are in fact the names of response-dependent properties, and if an account of morality that involves only these normative properties can manage to fulfill defensible adequacy conditions on a moral theory, then by the same token it will turn out that the correct moral theory was broadly consequentialist all along. Moral theory is more of a conceptual inquiry than an empirical one, and when one is providing conceptual analyses, it is not always a virtue of one's methodology that it leaves conceptual space for false accounts – for such accounts might well have been conceptual confused. This does not mean that one has to rule out deontological theories from the start, before beginning to pursue the line of theorizing I advocate. But it does mean that if one has pursued that line to a successful conclusion, then one shouldn't be bothered by the fact that one's method could not have yielded a radically different sort of view.

NOTES

☆ Thanks to Mark Timmons for helpful and friendly comments on an earlier version of this paper, and also to an audience at the 2003 Pacific APA, and to the reviewers for this journal.

[1] See Terry Horgan and Mark Timmons: 1991, 'New Wave Moral Realism Meets Moral Twin Earth', in John Heil and M. D Lanham (eds.), *Rationality, Morality, and Self-Interest*, Rowman and Littlefield, pp. 115–133. Other relevant papers by Horgan and Timmons include: 'Troubles for New Wave Moral Semantics: The "Open Question Argument" Revived', *Philosophical Papers* **21**, 153–175, 1992; 'Troubles on Moral Twin Earth: Moral Queerness Revived', *Synthese* **92**, 221–260, 1992; 'From Moral Realism to Moral Relativism in One Easy Step', *Critica* 28, 3–39, 1996.

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

[2] Sean Holland: 2001, 'Dispositional Theories of Value Meet Moral Twin Earth', *American Philosophical Quarterly* **38**, (2) 177–195.

[3] Paul Bloomfield has very recently offered another response to Horgan and Timmons. See his 'The Rules of Goodness: an Essay on Moral Semantic', *American Philosophical Quarterly* **40**, (3) 197–213, 2003. (Bloomfield) p. 202 also draws attention to the importance, in undermining the force of Horgan and Timmons's argument against the realist, of the *theoretical* nature of the moral terms. However, he does not seem to recognize the possibility of the sort of *basic* non-moral normative terms with which this paper is concerned. Rather, he focuses on an analogy between the complex property of moral goodness and the theoretical property of healthiness. I believe that Bloomfield is right in thinking that there is an important analogy here. But I also think that basic normative terms such as 'harm' and 'benefit' might be among the building blocks of an adequate account of healthiness, so that Bloomfield is too pessimistic in claiming that 'a straightforward definition [of 'healthiness'] is impossible', p. 204.

[4] It is the work of an advocate of this sort of causal theory of meaning to shed light on the particular nature of this 'certain causal role'.

[5] See Hilary Putnam: 1979, 'The Meaning of "Meaning,"' in his *Mind, Language and Reality*: *Philosophical Papers*, Vol. 2, Cambridge: Cambridge University Press, pp. 215–271; 'Meaning and Reference', *Journal of Philosophy* **70**, 699–711, 1973. For one criticism of this view of the meaning of 'water', see Davies and Humberstone, p. 19. But the point here is not the correctness of the view. Rather, it is whether the same sort of view is any worse when applied to 'morally good'.

[6] To make this point clearer by contrast, it may be worth noting that indexical words such as 'I' and 'here' are referring terms that do not tie meaning and referent so closely. So too are moral terms, according to a certain sort of cultural relativist.

[7] Compare Holland, p. 185.

[8] See Crispin Wright, 'Moral Values, Projection and Secondary Qualities', *Aristotelian Society Supplement* **62**, 1–26, 1988.

[9] This claim even accounts for the vagueness in 'red' by correlating it with the vagueness in 'very high'.

[10] Compare Hume, *An Enquiry Concerning the Principles of Morals*, p. 293, and Arthur Ripstein, 'Preference', in Christopher W. Morris and Arthur Ripstein (eds), *Practical Rationality and Preference: Essays for David Gauthier*, (New York: Cambridge University Press, 2001), p. 52. Ripstein laments Hume's final step from the avoidance of pain being an "ultimate end" to an apparently post-ultimate justification of its status. The current paper allows for a more charitable interpretation of Hume as offering a genealogy, and not a justification.

[11] Sometimes we call things harms because they involve a preponderance of harms over benefits. This raises important questions about how to balance harms against benefits and against each other. Such questions are beyond the scope of the present paper, which seeks only to show that a response-dependent account of 'harm' can avoid the problems that afflict a direct response-dependent account of more complicated normative notions. I address these problems in Brute Rationality: Normativity and Human Action, Cambridge: Cambridge University Press, 2004, esp. chap. 7.

[12] The defense of Dan White for the murder of Harvey Milk also suggests that one drawback of eating Twinkies may be that they induce a homicidal mania.

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions

[13] Compare S. L. Hurley, 'Objectivity and Disagreement', in Ted Honderich (ed), *Morality and Objectivity: A Tribute to J. L. Mackie*, (Boston: Routledge and Kegan Paul, 1985), pp. 54–97, at p. 74.
[14] Compare Ludwig Wittgenstein, *Remarks on the Foundations of Mathematics*, G. H. von Wright, R. Rhees, G. E. M. Anscombe, (eds), Volume III, (Oxford: Blackwell, 1978), p. 37.
[15] It is also helpful to keep in mind, during all discussions of how to draw the distinction between substantive disagreement and conceptual disagreement, that the distinction admits of degrees. See Hurley (1985), p. 80.
[16] Charles Stevenson, 'The Emotive Meaning of Ethical Terms,' *Mind* **46** 14–31, 1937.
[17] Some philosophers of course will object to an account according to which a rational person could possibly be indifferent to a reason of which he is aware. A full response to this interesting philosophical illusion is beyond the scope of this paper, but is addressed in my 'Practical Rationality, Morality, and Purely Justificatory Reasons', American Philosophical Quarterly 37, 227–243, 2000, 'Skepticism about Practical Reasons Internalism', Southern Journal of Philosophy 39, 59–77, 2001, and 'Korsgaard's Private-Reasons Argument', Philosophy and Phenomenological Research 64, 303–324, 2002.
[18] Such an account has much in common with the moral view defended by Bernard Gert, in his *Morality: Its Nature and Justification*, (New York: Oxford University Press, 1998).
[19] For some illuminating remarks in this connection, as well as a provisional definition of 'impartiality', See J.S. Mill, *Utilitarianism* (1861), George Sher (ed.), (Indianapolis: Hackett, 1979), pp. 44–5.
[20] This illustrative account of impartiality is also taken from the moral theory of Bernard Gert.

Department of Philosophy
The Florida State University
Tallahassee, FL 32306
U.S.A.
E-mail: jgert@fsu.edu

This content downloaded from 128.163.2.206 on Sun, 16 Feb 2014 20:17:09 PM
All use subject to JSTOR Terms and Conditions