

D. LOEB

MORAL REALISM AND THE ARGUMENT
FROM DISAGREEMENT

(Received in final form 26 June 1996)

The argument from disagreement is among the most important that has been directed against moral realism.¹ Yet, in recent years it has been widely rejected – so widely, in fact, that even those who are sympathetic with irrealism have often found it unconvincing. One reason for this has to do with the fact that newer, more sophisticated versions of moral realism have recently emerged. Some of these seem more resistant to the argument, in part because of the way they take advantage of philosophical resources not available to their predecessors – resources which have themselves emerged recently in philosophy of science, language, mind, metaphysics, and epistemology.

In this paper, I want to take another look at the argument from disagreement. I begin with a brief discussion of J. L. Mackie's version of the argument, which criticizes the epistemology of moral realism, and especially of moral intuitionism. Intuitionism, of course, is out of fashion now, but an examination of the argument in its earlier form will prove useful in our discussion of a more contemporary version of moral realism. Next, I turn my attention to one of the most promising of these, a prominent form of scientific naturalism, and try to explain why the argument still seems so dissatisfying to people on both sides of the debate, why it may be stronger than it has seemed, and what we would have to do in order to see just how strong it is. In the final section of the paper, I consider a less familiar version of the argument – one that focuses on the semantics of moral realism rather than on its epistemology. Here again, I conclude that the argument has more going for it than meets the eye, and that more work is needed to see whether it will ultimately succeed in undermining moral realism.

Philosophical Studies 90: 281–303, 1998.

© 1998 Kluwer Academic Publishers. Printed in the Netherlands.

I. INTUITIONISM AND THE ARGUMENT FROM DISAGREEMENT

The argument from disagreement holds that widespread differences of opinion on moral questions are evidence against moral realism. J. L. Mackie presented one of the most sophisticated treatments of the argument, directing it primarily at ethical intuitionism, which held that people have the ability to apprehend certain moral truths directly. According to Mackie, the best explanation for widespread differences in moral belief is not that our intuitive faculty often goes awry, but that these differences stem from pre-existing cultural differences:

In short, the argument from relativity has some force simply because the actual variations in the moral codes are more readily explained by the hypothesis that they reflect ways of life than by the hypothesis that they express perceptions, most of them seriously inadequate and badly distorted, of objective values. (Mackie, 1977, p. 37)

Mackie's hypothesis, then, was that we pick up our values from our culture and not from the operation of some special intuitive capability. But why should we think that Mackie's hypothesis is better? One way of filling out the argument is as follows: The plausibility of moral realism depends on the assumption that we have direct (non-inferential) access to morality. But moral disagreement shows that we do not have direct access to morality. So moral realism is implausible.

Each of these premises is open to challenge, of course. One might deny that the plausibility of moral realism depends on the assumption about direct access. But intuitionists did not want to deny this. Moral disagreement is also hard to deny. We see plenty of it within our own culture and even more across cultures and over time. To resist the argument from disagreement, then, the intuitionist must examine the connection between disagreement and direct access. Is it true that moral disagreement is evidence against our having direct access to morality?

The answer depends in part on how much moral disagreement there is. But it also depends on the way in which our access to morality is supposed to work. Suppose the intuitionist claims that we have the ability just to look at a situation and "see" its moral properties. The fact that we often see differently seems to be good evidence against this claim.

Even here, of course, the evidence is rebuttable. Take the analogous case of vision. Some people are blind or visually impaired in certain ways, and anyway, we see things from different perspectives and consequently see them differently. In fact, far from being inconsistent with our everyday hypothesis about vision, these cases appear to support it, since our theory predicts that they will occur.

The same may be said with respect to morality. Some people, it might be thought, are morally blind or have impaired moral sensibilities. Others are unable to see the truth owing to accidents of perspective, as when personal interest or social ideology skews a person's perception. In particular, our interest in seeing ourselves (and those with whom we identify) as good people may cause us to see our own behavior as morally acceptable, regardless of whether it is in fact acceptable. Thus, moral disagreement is compatible with the claim that we have sufficient access to the moral truth to make our belief in morality reasonable.

Of course only a few intuitionists ever actually held that intuition works at the level of individual cases, and Mackie was well aware of that fact. More commonly they held that we intuit only very general moral principles and their weights (such as that it is at least *prima facie* wrong to lie, and worse to kill), formal constraints on moral thinking (such as universalizability) or the nature of the good. And access of this more modest variety seems adequate to support the claim that morality is real. If this is what the intuitionists had in mind, Mackie acknowledged, then some apparent disagreement about moral questions could merely reflect differences in the circumstances to which these very general principles are *applied*. Thus, in a famous example, parricide may sometimes be thought morally permissible, but only, perhaps, where "the food supply is short and a choice has to be made between the aged and the young and healthy."² Had we found ourselves in similar circumstances, the argument goes, we might well have made similar moral judgments, because we accept the same fundamental moral principles.

A related point is that some apparent moral disagreements merely reflect differences in belief about the non-moral facts, often facts about the consequences of various contemplated actions. Even if we could intuit that pleasure is the only good, for example, we might still disagree about whether "ending welfare as we know it" by limiting

eligibility to two years would produce more or less overall pleasure, and thus disagree about the desirability of changing our policy.

How successful are these responses? To a large degree that is an empirical question. We already possess a great deal of information about the extent of moral disagreement, but we do not know as much about how well that information can be squared with the explanations offered by intuitionists. Typical intuitionist treatments of the argument merely set forth a few prominent examples of apparent moral disagreement, and pointed out that the disagreement appears to dissolve upon closer inspection. But it is not obvious that these examples can be generalized to all cases in which intuition is thought to operate. In exactly which cases does the best explanation for moral disagreement imply that the disagreement would evaporate if underlying non-moral differences were resolved? How often is apparent disagreement explicable as the application of identical principles in different contexts? More generally, how much real moral disagreement remains after all the sources of confusion discussed above have been removed? Resolving these questions without allowing the explanatory tail to wag the evidential dog would require philosophical imagination, but it would also require a great deal of further empirical research into the circumstances and beliefs of various cultures.

Irrealists cannot just wave their hands here either, of course. Even if claims about the extent of moral agreement are sometimes exaggerated, what agreement there is needs to be explained. Presumably most irrealists believe something like this: We are genetically similar to one another, and thus have many aims in common. More importantly, where our aims differ we still have a common interest in preserving a system that makes their realization possible, and doing so involves a good deal of compromise, as Hobbes and others have long pointed out. Certain general rules or principles, especially the more widely accepted ones such as the prohibitions on murder and theft, seem especially well-suited to the realization of these shared aims, and it is not surprising that people have learned this. Furthermore it would not be surprising if evolution had resulted in certain altruistic or cooperative tendencies becoming part of our genetic constitution, in virtue of the clear survival advantage for genetic lines which have them.

Still, this is only a hint of a reply, and as before much more would need to be done to show that it is correct. And once again the questions get complicated quickly, in part because realists can incorporate much of the irrealist's picture into their own explanatory story. These factors help to explain why the argument from disagreement has seemed so dissatisfying to both sides of the debate. It does not seem possible to form a reasonable judgment about who is right without an extensive treatment of a number of issues, some of them quite difficult, and many of them unlikely to be addressed soon. If I am right, the argument's fall from favor can be attributed more to stalemate than to defeat. But there is a further reason for the argument's fall from favor. Newer versions of moral realism have developed a variety of heretofore unavailable strategies for responding to it. In the remainder of this paper, I focus on one of the most prominent of these new forms of moral realism.

II. NATURALISTIC MORAL REALISM AND THE PROBLEM OF EPISTEMIC ACCESS

In recent years a new approach has replaced intuitionism as the dominant form of moral realism in America. I'll call this approach *Naturalistic Moral Realism*, or NMR, for short.³ Naturalistic Moral Realism shuns many of the extravagances of intuitionism and makes use of recent advances in other areas of philosophy to avoid some of the most serious arguments directed against moral realism in the past. Gone are the special metaphysical status for moral facts (complete with their own non-natural ontological realm) and the special faculty of intuitive access, which seemed necessary for us to apprehend such unusual facts. Moral facts are natural facts like any others, knowable by ordinary modes of inquiry such as observation and theory testing.

Of particular interest to us is the epistemological dimension of this transformation. If it is no longer claimed that we have a faculty for non-inferentially apprehending the moral truth, then there is no longer any point in calling that faculty into question. Still, Naturalistic Moral Realists continue to maintain that we have some form of access to morality. Morality could be real but inaccessible, of course, but if so then we would lack the best sort of evidence for believing in it. In fact, access is sometimes built into the very definition of

realism (see, e.g., Boyd, 1988, p. 181). So disagreement is still a cause for worry. Thus, some Naturalistic Moral Realists have interpreted Mackie's argument in roughly the following way:

- P1 If moral realism is correct, then moral questions must be, in principle, resolvable.
- P2 A number of moral questions are not, even in principle, resolvable (as evidenced by moral disagreement).
-
- C Moral realism is not correct. (See, e.g., Brink, 1984, p. 424.)

This version of the argument does not assume, as both the intuitionists and Mackie did, that if there is epistemic access to morality, then it must be non-inferential.

David Brink presents an important reply to this version of the argument, in part by challenging P2. As we saw earlier, some moral disagreements reduce to disagreements over non-moral facts. But, Brink claims, (optimistically, I think) these non-moral disputes are always resolvable in principle. (Brink, 1984, p. 424–25)⁴ Furthermore, newer epistemological approaches can help us to see how even purely moral disputes, which might once have been thought irresolvable, can be resolved after all. Mackie, according to Brink, assumed that genuine cases of moral disagreement can be resolved only if there is *antecedent* agreement on general moral principles. This fits Mackie's emphasis on intuitionist versions of moral realism, if (as Brink claims elsewhere) intuitionism is a form of foundationalism. According to moral foundationalism our justified moral beliefs are either self-evident (and hence non-inferentially justified) or justified by linear (non-looping) inferences from these foundational moral principles. If the foundational moral principles must be highly general, as most intuitionists held, then disagreement over them cannot be resolved by appeal to more specific moral judgments and principles.

But, argues Brink, "This claim presupposes a one-way view of moral justification and argument according to which moral principles justify particular moral judgments but not vice versa" (Brink, 1984, p. 424). Brink rejects this approach to moral justification, in favor of a coherentist model along the lines of Rawls's method of

reflective equilibrium. On this approach, conflicts between more and less general judgments and principles can be resolved in either direction, offering hope for the resolution of disputes even among highly general principles. Instead of claiming that those with whom we disagree about such principles are morally blind or that their intuitive capabilities have been warped by self interest, we can *reason* with them by pointing out the implausible implications their principles would have for questions about what it is right to do (or about what is good, etc.) in specific cases.⁵

Brink also challenges P1, arguing that certain types of irresolvable moral disputes do not threaten moral realism. In some cases competing moral considerations are of equal weight; in others, conflicting values are incommensurable, but nevertheless real. Neither sort of irresolvable conflict presents a problem for moral realism, Brink thinks, and thus “the moral realist need only maintain that *most* genuine moral disputes are resolvable” (Brink, 1984, p. 424, emphasis added). But these sorts of cases do not really support Brink’s conclusion. Cases involving equally weighted moral considerations *do* have correct resolutions – for example that both courses of action would be equally right. If one or both parties does not recognize that this is so, then there is still genuine moral disagreement of the sort that threatens moral realism. Similarly, in cases involving conflicting but incommensurable values, it should be possible, at least in principle, for each party to see the correctness of the other party’s position. Thus, although there may not be a single correct answer in these cases, there need not be any disagreement, since it should be possible for both parties to recognize that fact.

Another possible challenge to P1 is suggested by the recognition that a certain amount of indeterminacy is compatible with realism about an underlying subject matter. It is now widely allowed that terms referring to ordinary objects often have indeterminate extensions in borderline cases. (Is a teepee a house, for example?) If this is also true for moral terms, then perhaps the irresolvability of some moral conflicts can be explained as stemming from this indeterminacy. For example, we should not be surprised if there is no bright line separating persons (beings with a right to life) from non-persons. Yet a lot can hang on the distinction, and it is easy to see how people might disagree vigorously in particular cases. More gener-

ally, if moral properties have indeterminate borders, then disputes about borderline cases are to be expected and are compatible with moral properties being real.

Peter Railton offers a further challenge to P1, adopting what Mackie (perhaps uncharitably) called the “companions in guilt” approach: Even on matters of non-moral fact, Railton claims, we are nowhere near to possessing uncontroversial “canons of induction so powerful that experience would, in the limit, produce convergence on matters of fact among all epistemic agents, no matter what their starting points,” and there is considerable doubt that such canons even exist (Railton, 1986b p. 6.) But, although Railton is right to point out that the standard implicit in Brink’s formulation is too high, enough disagreement of the right sort *would* undermine the claim that we have access to morality, and with it the plausibility of moral realism.

NMR, however, has more to say about why we should expect disagreement about moral properties. First, if NMR is right that we do not get moral knowledge from some perception-like faculty of intuition, then perhaps we should think of our access to moral properties as very significantly *theoretical*, as in the case of physicists’ access to microphysical properties. And, just as we should expect more disagreement about the microphysical properties of an object than about its color and shape, we should expect more disagreement about moral properties if our access to them is significantly theoretical.

Actually NMR need not completely reject the idea that moral properties are observed in order to make this point, and it would not be plausible to do so. We do think of ourselves as making moral observations (in more than a metaphorical sense), and there is no need to abandon this thought. For according to much recent philosophy of science, *all* observation is theory-laden. Thus, theoretical considerations will play a role in moral observations, just as they do in any others. And that being the case, the core of this reply is still available to NMR: Differences of belief among moral reasoners should be expected because the same information will be observed differently depending on what background theories are present. And, as before, where theoretical concerns play a more significant role, we should expect even more disagreement, at least for the time being.⁶

NMR has yet another reply at its disposal. One way in which observation is theory laden has to do with certain social dimensions of knowledge. In particular, if moral knowledge is not gained by individual cognizers directly apprehending moral truths, then we can see that the transmission of moral knowledge is in large part a social undertaking. Consider, by analogy, most people's scientific knowledge. We get it from textbooks, magazine articles, science teachers, and other "expert" sources. We take their word for it, yet much of what we believe on that basis is knowledge just the same. In fact, even scientists' own knowledge is to a very large degree social in this way.

Moral knowledge, like scientific knowledge, is social in this sense. We learn it from others – from our parents, teachers, and religious leaders, from books and plays and television, and from others with whom we come into contact throughout our lives. Even our moral "observations" have a significant social component. Just as the scientist sees what she sees in part because she has been trained to, so we make the moral observations we do in part because of our moral upbringing.

The interesting point, for our purposes, is that the claim that moral knowledge is passed along socially is precisely Mackie's explanatory hypothesis. Far from being at odds with Mackie's explanation, NMR actually predicts that moral beliefs will be transmitted through culture. Thus Mackie's version of the argument from disagreement cannot be successfully directed at NMR.⁷ Furthermore, given the social transmission of moral knowledge, we should not be surprised if there is widespread disagreement on moral questions. Mistakes can be made in the transmission of moral beliefs or background theories from person to person. Likewise, the people doing the transmitting may themselves be partially mistaken, especially if, as has sometimes been suggested, moral theory is at a very early stage of development.

NMR has assembled an impressive array of new tools for replying to the argument from disagreement. Still, these tools do not represent a complete defense against the argument. As things stand, they are merely promissory notes, whose satisfaction will once again depend on further developments in philosophical anthropology and normative ethics. Some Naturalistic Moral Realists are very optimistic about the outcome of such inquiry. Boyd, for example, believes that

“careful philosophical examination will reveal . . . that agreement on nonmoral issues would eliminate *almost all* disagreement about the sorts of moral issues which arise in ordinary moral practice” (Boyd, 1988, p. 123).⁸ But conjecture about what careful examination will reveal is no substitute for careful examination, and the later is what is needed if we are to make further progress here.

Moreover, there is reason to doubt that things will go as well as Boyd imagines. As an example, consider the debate between Kantians and utilitarians. It seems very unlikely that the continued existence of this debate hinges upon disagreement over the non-moral facts. Presumably the two sides agree on many of those, and disagreement, where it exists, does not neatly track disagreement in moral theory nor explain much of the difference in moral belief between the two sides. Boyd, of course, admits that “for some few real-world cases and for *lots* of the contrived cases so prevalent in the philosophical literature there does appear to be serious difficulty in finding rational resolutions.” But, he says, “In such cases the strategy open to the moral realist *is* to insist that failures of bivalence do occur just as a homeostatic consequentialist moral realist predicts” (Boyd, 1988, p. 213).⁹ Applied to the debate between Kantians and utilitarians, the idea seems to be that the two theories pick out many of the same acts as morally right and wrong, and that what disagreement remains involves cases that are on the borderline between right and wrong – cases in which our moral questions have no determinate answers (Boyd, 1988, pp. 196–199, 200, 212–214).

But an account of morality that allowed for a great deal of indeterminacy would be both disappointing and suspect. It would be disappointing because an important goal of moral philosophy is to find answers to moral controversies about which there is significant disagreement. And it would be suspect because it seems plausible that, whatever its underlying nature, morality has answers (difficult though they may be to discover) to most of these controversies. For one thing, NMR itself draws a strong analogy between morality and science. Although Boyd and others have maintained that fuzzy borders are likely to exist in at least some scientific domains, such as the social sciences and parts of biology, there is a good deal of pressure against indeterminacy as well: Unless *something* holds putative scientific properties together fairly tightly they cannot count

as natural kinds. The same, presumably, is true of moral properties: If they are real, there had better not be too much indeterminacy about their extension.

Furthermore, where there is indeterminacy it should be possible, at least in principle, to recognize it as such, just as it should be possible to recognize the existence of equally weighted or incommensurable moral values. If we continue to disagree a great deal about borderline cases, rather than recognize that they are irresolvable *because* they are indeterminate, then we still have genuine disagreement of the sort that threatens moral realism.

The *source* of the alleged indeterminacy is also relevant to the plausibility of this sort of response. Typically we find indeterminacy in cases in which there is a continuum along which a certain factor (or family of factors) is present to a greater or lesser degree (as in the case of baldness). But not all cases in which the two theories disagree fit neatly into that sort of framework. A utilitarian, for example, might feel that it is permissible for a government to torture a terrorist's child in order to determine the location of a bomb it has reason to believe will be detonated soon. A Kantian might think that this impermissibly treats the child as a mere means. But it may be that neither side finds this a particularly close call, such that if more lives were at stake, for example, the Kantian would relent. She is not *counting* lives, but looking to considerations of an entirely different sort.

At this point the proponent of NMR may try a different tack, claiming that what is responsible for the disagreement in this case is a more fundamental disagreement in background theory, and reminding us that disagreements of this sort are predicted on his model. But it would be illegitimate to assume that the disagreement in cases like this typically *stems* from disagreement about theory. What makes these theories appealing in the first place may have more to do with the way they answer these more specific questions than with considerations of the more overarching variety. Thus, Kantianism may be more appealing to Kantians precisely because it gives the answers it does in cases in which it conflicts with utilitarianism.

On the other hand, the fact that the two moral theories offer very different explanations for the rightness or wrongness of various action is itself a source of vast disagreement. We *don't* just disagree

about a few borderline cases, such as whether it is impermissible to torture in a case like the one described above, but about *why* it is wrong to torture in a wide variety of cases. Focusing on the narrow question of what a person ought to do, without looking further, masks tremendous disagreement of this sort. But agreement in the *extension* of moral terms is insufficient for moral realism. Eventually, we need to agree on a single correct explanation for the rightness of various actions, and the explanation cannot be that either they maximize utility or treat others as ends. Yet we are nowhere near a satisfactory resolution of these theoretical issues.

NMR has still other tools at its disposal, of course, but these seem equally unpromising. Is it really plausible to think that one side of the debate is blinded by personal interest? Can we really be confident that a foundationalist approach to moral justification is to blame for our lack of progress in resolving fundamental questions in normative ethics, and that a more thoroughgoing application of the coherentist methodology will make one side of the debate a clear winner? If not, then our position is similar to the one we found ourselves in at the end of the previous section. No doubt both sides have hunches about how things will turn out. But at this point it seems irresponsible to make any pronouncements one way or the other.¹⁰

In what follows, therefore, I want to leave this more traditional version of the argument from disagreement aside, and look at a different worry raised by widespread moral disagreement – one that does not focus on access, but instead questions the *semantics* of moral realism: Given so many apparent differences of belief over moral questions, why think that we are all talking about the same thing?

III. THE SEMANTIC ARGUMENT

I turn now to a less familiar problem raised by moral disagreement. If people have widely differing beliefs about a number of moral questions, perhaps charity requires that we interpret them as referring to *different* properties, or using moral language in some entirely different, non-referring way (as non-cognitivists contend). That way, at least, we could avoid holding that people are so often in error about what morality requires. In this section, I explore this semantic version

of the argument from disagreement, and consider what appears to be NMR's best line of defense against it.

The semantic problem would be particularly acute if we were to adopt a traditional approach to meaning, according to which meaning is "in the head" – a function of individual belief. As Boyd points out, the traditional approach "entails that major changes in scientific theories are almost always to be diagnosed as changes in subject matter or conceptual framework, rather than as new discoveries." (Boyd, 1979, pp. 374). Thus, the proponent of the traditional approach is forced to say that scientists using the word 'mass' at the turn of the century were not talking about the same thing as contemporary scientists using the word, since they had significantly different beliefs about the referent of that term. Likewise, he is forced to say that if people have significantly different beliefs about morality, then they have different concepts of morality. Thus, if my beliefs involving the word 'goodness' differ significantly from those of a seventeenth century Puritan minister, then we are simply not talking about the same thing. Kuhn, of course, embraced this result for scientific terms, but Boyd and others have thought it a *reductio* of the traditional view of language.

The causal theory of reference, introduced by Kripke and Putnam, but adopted and refined by proponents of NMR, was developed partly in response to this sort of consideration. According to the causal theory, reference is not fixed by beliefs alone, but by an appropriate causal relationship between people's use of a term and an object or property in the world. Instead of referring to whatever it is that satisfies some pre-conceived definition, words refer to whatever entities or properties actually causally regulate their use. Definitions are thus synthetic and not analytic, naturalistic and not conventional. They are intended to set forth the fundamental nature of the subject matter under investigation, as discovered by *a posteriori* empirical inquiry.

The causal approach seems to offer a way out of the semantic predicament posed by the traditional view of language, by explaining how co-reference could take place in cases like those described above. Since reference is fixed by actual causal regulation, two or more communities (or perhaps individuals) are co-referring just in case their uses of the referring terms are causally regulated by the

same real entities or properties. Thus, whether co-reference is taking place is to some degree independent of what a person believes. For example, both Aristotle and contemporary chemists can refer to water even though many of their beliefs about it diverge sharply, since both are regulated in their use of 'water' (or '*hudor*') by water itself.¹¹

Similarly, Boyd suggests, the causal theory might allow NMR to answer the semantic worry raised by the traditional view of language:

Some philosophical opportunities are too good to pass up . . . For example, against the objection that wide divergence of moral concepts or opinions between traditions or cultures indicates that, at best, a constructivist analysis of morals is possible, the moral realist might reply that differences in conception or in working definitions need not indicate the absence of shared causally fixed referents for moral terms. (Boyd, 1988, p. 199)

Just as scientists of different eras can refer to the same physical entities, in spite of important differences of belief, so people of different cultures and backgrounds can refer to the same moral properties, in spite of wide-ranging differences in their moral opinions.¹² Moral terms such as 'good' and 'right' refer if and only if people's use of these terms is casually regulated by real entities or properties. If it is, then even if we disagree with one another about the nature of goodness, or about which things are good, we are all still referring to goodness. Of course, the causal history linking our use of the term (or translated term) with the property must be plausible for the claim of co-reference itself to be plausible. But plausible causal histories need not be ones in which all users of the term share a common set of beliefs about the entities or properties for which the term stands.

Are differences of belief irrelevant according to the casual theory, then? They shouldn't be. Granted, differences in belief are *compatible* with identity of reference. But whether and to what degree they are evidence against the hypothesis of shared reference depends on how much disagreement there is and on the nature of that disagreement. To make this point clear, let's begin with a non-moral example. Imagine a conversation in which Kevin and Konstantin are discussing their common interest in horses. They talk about how useful horses are for getting around town, how expensive they are to take care of, how much each cherishes his own, how they rub them down each day, and how beautiful they are to look at – especially the wild mustangs both have recently seen out in the country. Still,

we would give up the claim that they are both referring to horses if it became clear that their other beliefs diverge in certain ways – if, for example, Konstantin says that his horse has a red steering wheel and bucket seats.

This example is not meant to challenge the causal theory of reference. Instead what it shows is that a person's beliefs can provide us with relevant evidence on the question of what that person is referring to. Konstantin's belief that "horses have bucket seats" gives us excellent evidence that he is not really referring to horses. Thus, beliefs that appear to conflict still pose a problem for NMR, in spite of NMR's rejection of the traditional theory of meaning. In some cases, at least, such beliefs give us evidence that two or more uses of a term are not co-referential.

Of course we could continue to maintain that Konstantin is referring to horses, by claiming that he simply has many false beliefs about them (that they can have bucket seats, for example). But the only reasonable hypothesis is that he is mistakenly using the word 'horse' to refer to cars. And, indeed, that may well be because, given his beliefs, it is clear that his use of the word is casually regulated by cars and not by horses. His beliefs alone do not determine what he is referring to, according to the causal theory. That depends on the causal relations (including, of course, some involving his beliefs) between his uses of the term and horses or cars. But his beliefs are highly relevant to whether we are warranted in thinking that he is referring to horses.

The horse case is unlikely to occur in real life, of course. Typically, there is very little real disagreement about horses. But things are quite different in the moral case, where we do find a great deal of unresolved disagreement. In view of this disagreement, there is reason for concern about the semantics of moral discourse in a way in which there is no reason for concern about the semantics of ordinary discourse about horses. Perhaps, as some realists imagine, most of this disagreement can be resolved. But in the absence of such a resolution, the causal theorist is faced with two possibilities: The first is that when we speak in moral terms we are casually regulated by, and hence referring to, the same moral properties, but that people are often mistaken (and sometimes quite seriously mistaken) about these properties. The second is that we are not all

referring to the same things when we use the moral vocabulary, either because different people are referring to different things or because (as non-cognitivists and error theorists hold) we are not referring at all.

Thus, a full defense of moral realism along the lines proposed by NMR requires some plausible telling of the relevant causal story. Indeed, it requires that the causal story offered be more plausible than any of the alternatives and that it explain how co-reference is achieved in spite of widespread differences of belief. It would, perhaps, be unreasonable to demand too much detail here; such detail would not be available even in some non-moral cases where identity of reference is undisputed. But a general explanation of how the causal regulation takes place *is* available in those non-moral cases, and should be available here as well if we are to take the co-reference claim seriously.

The problem is that the sorts of causal stories that seem most plausible in other contexts (in which we are confident of co-reference) do not seem very plausible at all in the moral case. Return, for a moment to the example of Aristotle and the contemporary chemist. Both men can refer to water, even though their beliefs about water differ substantially. And it is easy to see how this can be so on the causal theory. Their talk about water is causally regulated by water itself. They both bathe in it, drink it, and see it in their lakes and rivers, and their use of the terms is directly regulated by these interactions. Call this a case of *unproblematic perceptual detection*.

By contrast, there is no unproblematic perceptual detection in the moral case. As we have seen, the claim that moral observations are more heavily influenced by theoretical considerations was an important part of NMR's explanation of how disagreement can be compatible with epistemic access. Thus, in developing a causal account for the moral case, we need to look elsewhere.

Another mechanism that has been used to explain how co-reference can be compatible with disparities of belief, has to do with what Putnam called, "the division of linguistic labor" (Putnam, 1975, pp. 227–229). Like a nuclear physicist, I can refer to protons, even though I believe them to be tiny balls somehow stuck to one another and to other tiny balls of a different color. Presumably the physicist has different beliefs, but it is still reasonable to presume

that we are co-referring, because it is still plausible to hypothesize an appropriate causal connection between my usage and real protons, just as it is for hers. The connection, in my case, is *mediated* by the connection between what certain experts have said and protons. For example, the causal chain might go from me to my eleventh grade physics text (and other similar sources) and from there to a community of experts who have themselves made theory-mediated judgments on the basis of causal contact with their apparatus, which is in turn causally affected by the behavior of protons.

There is, however, no division of linguistic labor in ethics. No one acts as a causal intermediary in the ordinary person's use of moral terms. When an ordinary person talks about protons, she intends to talk about whatever it is that physicists are talking about when they use the word 'proton.' But when she calls an action wrong, it seems very unlikely that she intends to talk about whatever it is that her minister, rabbi, or (of all things) philosopher is talking about when using the word 'wrong.' Thus, we must once again look for a different analogy.

One possible analogy is the specialists' *own* causal connections with the entities or properties in question. Putnam gives the example of Archimedes' reference to gold. Imagine Archimedes looking at a set of rock specimens arrayed before him, all of which he believes to be gold. Let us assume that all of these specimens are shiny, malleable, and yellow, and that Archimedes believes these to be the defining characteristics of gold. Still, some of the samples are fool's gold (pyrite), and it would be a mistake, Putnam thinks, to hold that Archimedes is referring to these when he uses the word 'gold'. Archimedes does not use 'gold' to refer to whatever has the properties he believes definitive of gold, but rather to refer to whatever has the real properties of the majority of rocks normally thought of as gold. As Putnam says:

If we put philosophical prejudices aside, then I believe that we know perfectly well that no operational definition does provide a necessary and sufficient condition for the application of any such word. We may give an 'operational definition', or a cluster of properties, or whatever, but the intention is never to 'make the name synonymous with the definition'. Rather, 'we use the name *rigidly*' to refer to whatever things share the *nature* that things satisfying the definition normally possess. (Putnam, 1975, p. 238)

As in the previous case, *intentions* are crucial here: Archimedes intends to refer only to a certain natural kind – things that have the same underlying structure as most of the things currently classified as gold. This intention picks out the causal relations that fix the reference of ‘gold’ when he uses it. In this case, his intention is *deferential*. He wants his language to “cut nature at its joints,” and thus he would rather admit that he was wrong about which of these samples is gold than claim that he had just been talking about *any* shiny, malleable, yellow rock, no matter what its internal structure. Thus, his usage is responsive to the distinction that actually exists between gold and fool’s gold. And if this is true, then his use of the term ‘gold’ is causally regulated by the substance, gold, in spite of his false beliefs about it.

But is the same thing true in ethics? Are our intentions with respect to terms in the moral vocabulary deferential, as they would need to be for the causal theory to explain how co-reference is possible in spite of widespread moral disagreement? No doubt under certain circumstances we would be willing to say that we were mistaken about what actions are right or wrong – if, for example, we were shown that our beliefs about them are the products of misconceptions about the non-moral facts, or of bias. But there are limits to our willingness to accommodate our moral beliefs to a moral theory, even a theory that can pick out many of the paradigm case of right action.

Imagine that some version of utilitarianism could classify as morally right most of the actions we would judge to be right, and few that we would not so judge. Suppose further that a Kantian were made aware of this fact. It seems very unlikely that she would change her mind about the nature of rightness, for to do so would require that she abandon certain strongly held moral beliefs that are incompatible with utilitarianism, such as that it is wrong to torture the terrorist’s child. In fact, it seems likely that many ordinary people, unpossessed of any well-articulated moral theory, would not change their minds about cases like this either. If not, then they do not intend to refer to just any property whose extension closely overlaps the range of cases they think right.

No doubt, the conviction that we are all talking about the same thing is a powerful one. And presumably that conviction is reflected

in intentions that are at least somewhat deferential. But the question is *how* deferential, and the worry for moral realism is that our deference is insufficient to resolve conflicts like these. If so, then it is hard to see how NMR can use the causal theory to reply to the semantic argument. The paradigm cases for explaining co-reference under disagreement seem not to apply to moral terms.

Someone might raise a different objection here, however: How can an issue about moral language be important to the metaphysical debate over whether moral properties are real? Isn't reality independent of how we choose to use our words? The answer is yes and no. Our language has no bearing on the question of what properties or entities exist. What exists exists whether we talk about it or not, and there are undoubtedly many things for which we have no words. But our language *is* relevant to the question of whether certain existent properties and entities are the ones we have been talking about when using a certain vocabulary. Thus, moral rightness is whatever people using the word 'rightness' have been talking about. If those who use the term 'right' are not referring to the same thing, then there is not a single property of rightness after all. And, as Boyd himself acknowledges (in another context), this would be fatal to realism about the properties or entities in question:

It is, of course, possible for a term *T* to afford epistemic access to several quite different kinds. The term "demon" probably afforded epistemic access to a great variety of kinds of natural phenomena for centuries. What I am suggesting is that it is correct to talk of the referent of a general term precisely in those cases in which the term affords epistemic access to a single kind or, at any rate, to a family of closely related ones. (Boyd, 1979, p. 385)

Likewise, even if there were some unique natural property (or homeostatic property cluster) that regulates some significant group of people's use of the term 'right,' it would be a serious blow to NMR if it turned out that different properties (or clusters) regulate the use of this term in other significant groups. We would have to hold that there are many different senses of 'right,' and that the term does not have a single referent as posited by NMR. And, as in the case of 'demon,' this relativization of 'right' would make it implausible that there is a real property of moral rightness.¹³

It is, however, a significant virtue of this version of the argument from disagreement that it need not attack moral realism on the metaphysical ground that the properties realists describe are unreal.

For example, Peter Railton holds that “an individual’s good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality,” and he uses this account of non-moral goodness to build up his account of moral rightness (Railton, 1986b, p. 16; see also Railton 1986a). The irrealist need not deny that there are real properties conforming to Railton’s criteria.¹⁴ Even if there are, the irrealist can still ask whether they are the properties of non-moral goodness and of moral rightness, as Railton contends. The answer depends on what, if anything, we are actually referring to when we use those words (or, in NMR terms, what properties actually regulate that use).

There are, as I have noted, other possibilities, besides reference to multiple properties. Non-cognitivists believe that we do not use moral terms to refer at all, but to recommend, endorse, prescribe, or emotive. One advantage of their approach is that it allows for a sense in which all who use the term are using it in the same way, by giving up the claim that this use is referential. Error theorists like Mackie hold that we are all attempting to talk about real moral properties, but that there are no such properties, so we are in error when we do so. The irrealist need not choose between these various alternatives in order to defeat moral realism. All she need do is show that their disjunction is more plausible than the hypothesis of common reference. Moral disagreement presents strong evidence in favor of this claim.

What would it take to determine whether or not the hypothesis of common reference is mistaken? All of the questions that came up in the previous sections are relevant here as well: How much moral disagreement is there? How much can be explained as owing to differences over the non-moral facts? How much as owing to bias, or to the application of identical principles in different material contexts? But our intentions are also relevant here, and we will need to learn more about them if we are to successfully evaluate this semantic version of the argument. Still, when we think about the contrast between what we want our moral language to do and what we want our scientific language to do, it is clear that it will be much more difficult to reconcile (apparent) moral disagreement with common reference than it is in the case of scientific disagreement.

Thus the problem of moral disagreement is more serious than NMR has yet acknowledged.

IV. CONCLUSION

There is no doubt that the Naturalistic Moral Realists have resurrected moral realism in a more powerful form, and that in doing so they have also helped to reinvigorate the metaethical debate. In particular, moral realism now has many new resources with which to combat the argument from disagreement. But we should not let the availability of these resources lull us into rejecting either version of the argument too quickly – in one case with the vague hope that sufficient moral disagreement can be resolved to warrant the claim of access, and in the other with the vague hope that the causal theory can rescue the claim of co-reference. At this point, confidence that moral realism can survive these challenges would be premature.¹⁵

NOTES

¹ Mackie, 1977, called it the argument from *disagreement*.

² Brandt, 1959, pp. 94–95. See also, pp. 99–103. For a slightly earlier discussion of the argument from disagreement, and of this example in particular, see A. C. Ewing, 1953, pp. 111–118.

³ Naturalistic Moral Realism is sometimes referred to as “New Wave Moral Realism” or, because of the institutional affiliation (past or present) of several of its most prominent proponents, as “Cornell Realism.” See, for example, Boyd, 1988; Brink, 1984, 1989; Sturgeon, 1984; Miller, 1985. A possible exception is Peter Railton, 1986a, 1986b. NMR is not, of course, monolithic. Important differences exist between its various proponents, and one could argue that some of those often lumped together under this heading do not belong in the class. What I will be presenting is a composite of some of the most prominent views. It may be that not everyone in the class accepts all of the positions I attribute to NMR, but the differences should not have a substantial impact on my argument.

⁴ Brink himself questions this claim in a later work. Brink, 1989, p. 199.

⁵ For an interesting challenge to coherentist versions of moral realism, see Timmons, 1990.

⁶ If this line of thinking is correct, then the rejection of intuition is not what is doing the work after all, but rather the rejection of a sharp observation/theory distinction. Even the intuitionist could take advantage of that. That is, even if there were such a faculty, observations would differ along with background theories. (We have a faculty of vision, after all, but it is well known that theory can influence visual perception.) Still, although the degree to which a particular type of observation is theory-laden will depend on the circumstances, intuitionists seem

committed to holding that moral intuitions are very close to the observational end of the spectrum. If so, then they would not be in a position to claim the full advantage of this reply.

⁷ Once again, it would seem, even intuitionists could take partial advantage of these arguments. The social nature of knowledge is itself a continuum, and just as we must accept that there is a social component to even “pure” observations (like that there is a book on the table) intuitionists should accept that moral observations are to some degree social creations. Still, the idea that moral intuition is largely observational does seem to push the intuitionist towards the less social end of the spectrum. And, inasmuch as intuitionists want to say that moral observation is more like the book-on-the-table case (and less like the case of my knowledge of particle physics), they locate it at the less social end of the spectrum, and thus have less room to move away from the problems moral disagreement creates.

The points raised in this and the previous note suggest that the line between intuitionism and other forms of realism may be fuzzier than we have thought. Whether we use the word “intuition” or “observation” to refer to our ability to recognize moral truth seems unimportant, given a correct understanding of how such recognition takes place. Hilary Kornblith brought this line of thinking to my attention.

⁸ Sturgeon is significantly more guarded on this question (see, Sturgeon, 1985, p. 49).

⁹ See also, Shafer-Landau, 1994, 1995. Boyd does not say why he thinks the distinction between real-world and imaginary cases so significant.

¹⁰ The prospects may be worse when it comes to resolving cross-cultural differences. Confucianists, for example, have very different attitudes than Kantians *or* utilitarians about the fulfilment of social roles, the relationship between personal and social morality, and the importance of hierarchy. Yet once again it seems implausible to attribute these differences to differences in non-moral beliefs, bias, or borderline indeterminacy. For interesting comparisons of Chinese and Western ethics, see the introduction to Munro, 1985; Hansen, 1985.

¹¹ This example is discussed in Miller, 1985, pp. 514–517.

¹² Boyd’s own interpretation of the causal theory implies an important connection between the semantic worry and the more traditional worry about epistemic access, and may thus help to explain part of the appeal of P1, above (Boyd 1979). Boyd however, does not seem to draw this connection in Boyd, 1988.

¹³ If the significant differences were only cross-cultural (or cross-temporal), we might be tempted to say that *we* are talking about rightness and have simply mistranslated other cultures’ terms. But that makes rightness seem too parochial. For moral realism to have any bite, it seems to me, the moral properties must be those referred to by a wide variety of cultures that appear to be talking about morality.

¹⁴ I question whether there are such properties in Loeb, 1995.

¹⁵ I wish to thank my colleagues, Sin yee Chan, David Christensen, Arthur Kutlik, William Mann, and Derk Pereboom, for their patient reading of various drafts of this paper. Hilary Kornblith was especially effective in his role as NMR foil. I have learned a great deal from my conversations with him. Finally, Barbara Rachelson provided much more than a sympathetic ear and incisive commentary. I couldn’t have done this without her support.

REFERENCES

- Boyd, R.N. (1979): 'Metaphor and Theory Change: What is 'Metaphor' a Metaphor for?', in A. Orotney (ed.), *Metaphor and Thought*, Cambridge: Cambridge University Press.
- Boyd, R.N. (1988): 'How to be a Moral Realist', in Sayre-McCord, pp. 181–228.
- Brandt, R.B. (1950): *Ethical Theory*, Englewood Cliffs: Prentice-Hall.
- Brink, D.O. (1984): 'Moral Realism and the Skeptical Arguments from Disagreement and Queerness', *Australasian Journal of Philosophy* 62, 111–125; reprinted in Louis Pojman, *Ethical Theory*. Belmont, Wadsworth Publishing Co, 1989. (Page references are to Pojman.)
- Brink, D.O. (1989): *Moral Realism and the Foundations of Ethics*, Cambridge: Cambridge University Press.
- Ewing, A.C. (1953): *Ethics*, New York: Free Press.
- Hansen, C. (1985): 'Punishment and Dignity in China', in D.J. Munro (ed.), *Individualism and Holism: Studies in Confucian and Taoist Values*, Ann Arbor: Center for Chinese Studies, The University of Michigan.
- Loeb, D. (1995): 'Full-Information Theories of Individual Good', *Social Theory and Practice* 21, 1–30.
- Mackie, J.L. (1977): *Ethics*, Harmondsworth: Penguin Books.
- Miller, R.W. (1985): 'Ways of Moral Learning', *Philosophical Review* 94, 507–556.
- Munro, D.J. (ed.) (1985): *Individualism and Holism: Studies in Confucian and Taoist Values*, Ann Arbor: Center for Chinese Studies, The University of Michigan.
- Putnam, H. (1975): 'The Meaning of 'Meaning' ', reprinted in *Mind, Language and Reality*, Cambridge: Cambridge University Press.
- Railton, P. (1986a): 'Moral Realism', *Philosophical Review* 9, 163–207.
- Railton, P. (1986b): 'Facts and Values', *Philosophical Topics* 14, 5–31.
- Sayre-McCord, G. (1988): *Essays on Moral Realism*, Ithaca: Cornell University Press.
- Shafer-Landau, R. (1994): 'Ethical Disagreement, Ethical Objectivism and Moral Indeterminacy', *Philosophy and Phenomenological Research* 54, 331–344.
- Shafer-Landau, R. (1995): 'Vagueness, Borderline Cases and Moral Realism', *American Philosophical Quarterly* 32, 83–96.
- Sturgeon, N. (1985): 'Moral Explanations', in D. Copp and D. Zimmerman (eds.), *Morality, Reason and Truth*, pp. 49–78, Totowa: Rowman and Allanheld; reprinted in Sayre-McCord, 1988.
- Timmons, M. (1990): 'On the Epistemic Status of Considered Moral Judgments', *Southern Journal of Philosophy* 29 Sup., 97–129.

Department of Philosophy
The University of Vermont
Burlington, Vermont 05401-3404
USA

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.