Moral Realism without Convergence

Sarah McGrath
Princeton University

ABSTRACT: It is sometimes claimed that if moral realism is true, then rational and informed individuals would not disagree about morality. According to this line of thought, the moral realist is committed to an extremely substantive convergence thesis, one that might very well turn out to be false. Although this idea has been accepted by prominent moral realists as well as by antirealists, I argue that we have no reason to think that it is true, and that the only convergence claims to which the realist is committed are trivial ones.

I. INTRODUCTION

As things stand, there is a great deal of persistent moral disagreement. But suppose that we idealize those who disagree in two respects, so that:

> (1) They are fully informed about any relevant nonmoral facts, and

> (2) They are fully rational.

How much moral disagreement would remain? One view is:

> **CONVERGENCE**: Rational individuals who were fully informed about the nonmoral facts would not hold incompatible moral views.

Whether CONVERGENCE is true is an interesting question about which much could be said, but it will not be my primary concern in what follows. Rather, the thesis with which I will be primarily concerned in this paper is the following:

> **REALISM REQUIRES CONVERGENCE**: If moral realism is true, then rational individuals who were fully informed about the nonmoral facts would not hold incompatible moral views.

As we will see, a number of prominent and influential philosophers have maintained that REALISM REQUIRES CONVERGENCE is true. My aim in this paper is to show that we lack good reasons to think that this is so.

II. CLARIFYING THE ISSUE AND THE STAKES

First, let's clarify the thesis of CONVERGENCE.

It is obvious that there is a great deal of actual moral disagreement. Less obvious is how *modally robust* that disagreement is, or the extent to which it would survive in certain counterfactual circumstances. As is often noted, at least some moral disagreement is due to disagreement about underlying nonmoral facts. For example, two people might agree that capital punishment is morally permissible if and only if it effectively deters murder, but nevertheless disagree about its permissibility in virtue of disagreeing about its effectiveness as a deterrent. In such cases, the persistence of the moral disagreement is contingent on the nonmoral ignorance of at least one side to the dispute.[1]

How much moral disagreement would remain if we had complete knowledge of all of the potentially relevant nonmoral facts? This question is controversial and inspires a wide range of answers.[2] But it is plausible that even against an assumed background of full information about the nonmoral facts, a significant amount of moral disagreement would persist so long as the moral views at which people arrive are unduly influenced by considerations of self-interest and other biasing or distorting factors. For example, even if the full effects of a proposed scheme of redistributive taxation were well known and vividly appreciated by everyone, there might still be significant disagreement about the justice of the proposal among members of different economic classes. The thesis of CONVERGENCE is consistent with disagreement that is due to the operation of such distorting factors: in considering whether fully informed and *rational* individuals would continue to disagree, we abstract away not only from disagreement based on ignorance but also from disagreement due to irrationality.

Next, let us clarify the position of the moral realist. For our purposes, we can understand moral realism broadly, as the conjunction of the following three theses:

> **Cognitivism**: Paradigmatic moral judgments are truth-apt.

Objectivity: Which moral judgments are true does not depend on what we (either individually or collectively) accept. Thus, subjectivist and relativist views about morality (e.g., Harman 2000) are false.

No Error Theory: At least some judgments that predicate moral properties are true. Thus, error theories of the kind put forth by Mackie (1977) and Joyce (2001) are false.

Understood in this broad way, moral realism is tantamount to the claim that (as it is sometimes put) "there are objective moral truths."

On this understanding of moral realism, it seems clear that moral realism does not logically *entail* CONVERGENCE. So it should be generally agreed that there are at least some possible versions of moral realism that are consistent with the falsity of CONVERGENCE. However, it is sometimes suggested that any *credible* form of moral realism incurs a commitment to CONVERGENCE. Interestingly, this idea has been endorsed by philosophers who differ widely in their attitudes toward moral realism itself. Thus, the idea has been endorsed by those who accept moral realism (e.g., Smith 1994, 2004, 2007), by those who reject it (e.g., Wright 1995), as well as by those who profess agnosticism about its truth. As an example of a philosopher in the last category, consider John Rawls at the time of his essay "The Independence of Moral Theory" (1974):

> It is natural to suppose that a necessary condition for objective moral truths is that there be sufficient agreement between the moral conceptions affirmed in wide reflective equilibrium, a state reached when people's moral convictions satisfy certain conditions of rationality. Whether this supposition is correct, and whether sufficient agreement obtains, we need not consider, since any such discussion would be premature. (290, cf. 301)

In his later writings, Rawls grew increasingly skeptical about the idea that we should expect all reasonable and informed individuals to converge in their substantive moral views, and he consistently eschewed talk of "objective moral truths."[3]

Indeed, the prospect that significant moral disagreement might persist even among reasonable and informed individuals raises the possibility of deploying REALISM REQUIRES CONVERGENCE in an argument against moral realism. Here, for example, is Crispin Wright:

> . . . it is of course evident that moral disagreements can be and frequently are attributable to confused thinking, factual ignorance and sheer prejudice. But the [realist's] obligation . . . is to show that such deficiency *has* to be involved in the generation of any such dispute . . . Any student of morality who has come to feel, therefore, that a substantial body of the principles that inform our ordinary moral thought are essentially contestable, and that no rational or cognitive deficiency is needed to sustain the clashes on things like sexual morality, the value of individual freedom, the moral status of animals and the ethics of suicide and mercy-killing, which are freely exemplified within and across cultures, won't give much for the realist's chances. ([1995] 2003, 199)

Consider then the following argument:

(P1)   *Realism requires convergence*: If moral realism is true, then CON-
VERGENCE is true.

(P2)   But CONVERGENCE is false.

(C)   Therefore, moral realism is false.

As the passage from Wright suggests, the second premise (P2) strikes many people
as quite plausible: they think that although eliminating nonmoral disagreement
and idealizing away from "sheer prejudice," "confused thinking," and other lapses
in rationality would eliminate at least some moral disagreement, significant moral
disagreement is likely to remain.

Against this, realists who accept REALISM REQUIRES CONVERGENCE
attempt to make CONVERGENCE plausible in various ways. For example, Smith
(1994), following Parfit (1984), appeals to the distorting influence of religious
authorities on moral thought.[4] Others emphasize just how much actual disagree-
ment might be explained in terms of the conflicting self-interests of those who dis-
agree, and by invoking similar mechanisms.[5]

However, there are reasons to be skeptical of attempts to defend realism by
defending CONVERGENCE in this way. While we shouldn't underestimate the
potential explanatory power of such accounts, there are significant limits to what
considerations of this kind might show. For it seems that considerations of this
kind could at best show that actual disagreement is relatively weak evidence that
CONVERGENCE is false, as opposed to providing reason to think that CONVER-
GENCE is true. Philosophers who appeal to the various ways in which nonmoral
ignorance and irrationality can underwrite moral disagreement emphasize the dis-
tance between people as they actually are and their idealized selves. But in this con-
text, the significant distance between people as they actually are and their idealized
selves is something of a double-edged sword. For given the distance between the
actual and the ideal, it seems as though the claim that CONVERGENCE *does* hold
will inevitably be extremely speculative—even if the claim that CONVERGENCE
does *not* hold is *also* quite speculative.[6]

A crucial point here is that even if we knew with certainty that (e.g.) consider-
ations of self-interest explain why two individuals believe as they do about distrib-
utive justice, and their views are incompatible, it does not follow (and might not
even be particularly likely that) the individuals would agree if they stopped giving
too much weight to self-interested considerations. More generally: Suppose that S1
and S2 hold incompatible moral views p1 and p2. It is tempting to conclude that,
if the explanation for why S1 believes p1 or the explanation for why S2 believes p2
involves nonmoral ignorance or irrationality, then S1 and S2 would agree in the
absence of the ignorance or irrationality. But that inference is fallacious: perhaps
they would end up converging if the nonmoral ignorance or irrationality were
removed, but perhaps they wouldn't.

An additional complication is the possibility that the elimination of nonmoral ignorance and irrational biases would *increase*, rather than decrease, the amount of disagreement with respect to at least some moral issues. This will be especially likely in cases in which the ignorance and biases are widely shared. For example, Singer (2005) speculates that certain near-universal moral intuitions are based on discredited systems of belief, and thus should be set aside when we engage in moral deliberation and theorizing. *Perhaps* doing this would ultimately lead to greater convergence in moral outlook. But on the face of it, it seems just as likely that deliberately setting aside widely shared moral intuitions—intuitions that might very well have approximated common ground if they were *not* set aside—might increase rather than decrease the diversity of the moral views at which people ultimately arrive. For example, perhaps there is a widespread moral consensus (among the population generally, even if not among academic philosophers) that it is at least morally permissible to bestow certain benefits on one's kin over arbitrary strangers who are objectively worse off. If the popularity of this view were revealed to be due to the operation of a rationally indefensible although widespread bias, this might very well increase, rather than decrease, the amount of disagreement about how we are morally required to bestow benefits on others.

In short, if the kinds of stories that philosophers tell in an attempt to minimize the significance of actual moral disagreement are the best that can be told, then it seems like we should not invest much credence in CONVERGENCE. (Even if such stories *do* succeed in showing that we should not invest much credence in the negation of CONVERGENCE.) If REALISM REQUIRES CONVERGENCE is true, we should not invest much credence in moral realism. (In effect, the credence that it is reasonable to give to the speculative claim CONVERGENCE sets an upper bound to the credence that it is reasonable to give to moral realism.) Thus, a moral realist who is committed to CONVERGENCE is giving a very large hostage to fortune.[7]

Again, my thesis is that there is no good reason for the realist to saddle herself with this heavy-duty commitment. Why might someone think otherwise? As we will see, there is no single reason why philosophers who endorse REALISM REQUIRES CONVERGENCE think that it is true. Rather, different philosophers offer different reasons for accepting it. In the sections that follow, I will examine a number of distinct motivations. These include considerations having to do with the putatively *a priori* character of morality (section 3), the aims of moral judgment (section 4), the alleged absurdity of "unknowable moral truths" (section 5), and the realist's commitment to the objectivity of morality (section 6). I will argue that none of these provides a good reason to accept REALISM REQUIRES CONVERGENCE.

Before turning to this task, a final point of clarification is in order. Consider the claim that:

> **REALISM REQUIRES CONVERGENCE ON THE TRUTH:** If moral realism is true, then rational individuals who were fully informed about the nonmoral facts would not hold incompatible moral views because they would converge on the *truth* about morality.

This thesis is logically stronger than REALISM REQUIRES CONVERGENCE. At least in principle then, a philosopher might clear-headedly accept REALISM REQUIRES CONVERGENCE while denying this logically stronger claim. That is, in principle a philosopher might endorse the following claim:

> If moral realism is true, then rational individuals who were fully informed about the nonmoral facts would not hold incompatible moral views even if they did not converge on the truth about morality.

However, philosophers who accept REALISM REQUIRES CONVERGENCE generally do so *because* they accept the logically stronger claim that REALISM REQUIRES CONVERGENCE ON THE TRUTH.[8] For this reason, I'll devote most of my attention in what follows to arguing that we lack good reasons to accept the logically stronger claim. In the final section of the paper, I will consider more briefly the question of whether we should accept REALISM REQUIRES CONVERGENCE even in the absence of good reasons for accepting REALISM REQUIRES CONVERGENCE ON THE TRUTH.[9]

III. A PRIORITY

The picture behind REALISM REQUIRES CONVERGENCE ON THE TRUTH seems to be this: if moral realism is true, then an inquiry conducted by individuals who are rational and well informed about the nonmoral facts will *inevitably* lead them to moral views that are true. Moreover, not only would such an inquiry not lead to any moral views that are false, but it would be *guaranteed* to deliver any genuine moral truths that there are.

On the face of it, the picture is a strange one, for the following reason: in nonmoral domains, realism is strongly associated with the thought that even idealized inquiry might fail to deliver the truth. For example, the realist about the past will allow for the possibility that some historical facts are *evidence transcendent*, in the sense that they would elude detection even by perfectly rational individuals employing our best methods for investigating the past. Indeed, in a possible world in which the empirical evidence that we have to go on is consistently misleading or unrepresentative, perfectly rational individuals will not only fail to hit on the truth but will be led further and further astray by their rationality. In fact, it has sometimes been taken as definitive of a realist stance about some domain that one thinks that truth in that domain is *not* epistemically constrained, that it might prove elusive even when pursued by perfectly rational people impeccably applying our best methods of inquiry; accounts of truth that tie the notion too closely to the deliverances of idealized inquiry are treated as paradigms of *anti*realism.[10] If we understand moral realism in parallel to realism about other domains, it seems like the moral realist should be the first to admit that in principle the moral facts could elude even the best efforts of idealized inquirers. But that things are parallel in this respect is often denied.[11] Why should morality be different?

Consider the following reply:

> Morality is different because of its *a priori* character. When it comes to empirical questions, even perfectly rational people will often fall short of believing the truth, for they are hostage to the quality of their evidence. Even if it is true that George Washington had an even number of hairs on his head at the moment of his birth, a perfectly rational individual will not believe this unless her evidence favors that possibility over the alternative. (Indeed, in such cases, her being fully rational partially explains why she does not believe what is true: an imperfectly rational individual who is not so constrained might come to believe the relevant truth on a whim.) And in cases in which the available evidence is misleading, a fully rational individual will believe what is false.
>
> But there are simply no analogous possibilities in the moral case. Knowledge of the most fundamental moral truths, like mathematical knowledge, can in principle be attained via the exercise of pure reason alone. Therefore, a fully rational human being would know the most fundamental truths of morality in this way. When it comes to ethics, there is no potential gap between the outcome of perfectly rational inquiry and the truth. Of course, not every moral truth is knowable a priori. For example, the knowledge that *it is wrong to encourage children to smoke* depends partially on empirical facts about the common effects of smoking on health. But the hypothetical individuals with whom we are concerned are not only rational but also know any relevant non-moral facts. Therefore, such individuals would know not only the most fundamental moral truths (a priori, in virtue of their rationality) but also the less fundamental moral truths, by combining their knowledge of the most fundamental moral truths with their knowledge of the relevant nonmoral facts. Therefore, any individual who is both fully rational and informed about the nonmoral facts would be *omniscient* about morality.
>
> But if any individual who is both fully rational and informed about the nonmoral facts is omniscient about morality, then no two individuals who satisfy these conditions will disagree about any moral question.

Although tempting, this line of thought is fallacious. First, even if morality is an a priori domain, in the sense that any fundamental moral truth that is knowable is knowable a priori, it doesn't follow that every fundamental moral truth is knowable. Compare: mathematics is an a priori domain if any domain is. Still, even if any mathematical truth that is knowable is in principle knowable a priori, it doesn't follow that every mathematical truth is knowable. Number theorists assume that Goldbach's Conjecture[12] is either true or false, despite the fact that they currently possess neither a counterexample nor a proof of it. Realists about mathematics allow for the possibility that Goldbach's Conjecture might be true even if no such proof exists, in which case Goldbach's Conjecture is an example of an unknowable mathematical truth. So even if mathematics is an a priori domain, it doesn't follow that someone who is fully rational will be omniscient with respect to mathematics, for some mathematical truths might lie outside the reach of pure reason. Similarly, even if morality is an a priori domain, it doesn't follow that someone who is fully rational will be omniscient with respect to the fundamental truths of morality.

Moreover, it is a mistake to suppose that a rational person will know—or even be in a position to know—any truth that is in principle knowable a priori. Rationality does not entail omniscience with respect to the a priori. Even if Goldbach's Conjecture is provable in some ingenious way (suppose that the proof ultimately comes to be regarded as one of the signature achievements of twenty-third-century mathematics), this does not mean that contemporary number theorists who are agnostic about its truth are guilty of any failure of rationality. On the contrary, their agnosticism seems to be the reasonable response to their epistemic situation.

Of course, one might simply use the term "rational" and its cognates in such a way that a "rational" person will be omniscient with respect to the truths of any a priori domain.[13] But that terminological choice would be ill advised. After all, paradigmatic failures of rationality include cases in which people arrive at unjustified conclusions by reasoning fallaciously, or fail to respect their evidence out of wishful thinking, dogmatism, or carelessness. Our failures to believe (e.g.) the most ingenious results of twenty-third-century mathematics, or truths that would be provable only via the use of supercomputers, have too little in common with these and other paradigms to make up an epistemologically useful category.[14]

But suppose that we did use 'rational' and its cognates in this inclusive way, so that full rationality entails omniscience with respect to a priori domains. In that case, given that morality is an a priori domain, it would follow immediately that rational individuals would not disagree about morality. However, it would be a mistake to think that convergence among such hyper-idealized individuals would do anything to defuse the challenge to moral realism that Wright and others mean to pose. In the passage quoted earlier, Wright raises the specter of continuing moral disagreement among individuals who are free from "confused thinking," "sheer prejudice," and "factual ignorance," and he suggests that disagreement of this kind is not something that the realist can countenance. Even if we agree to say that a fully rational individual would not only be free from relatively mundane defects of this kind, but would also be omniscient with respect to any a priori subject matter, Wright's challenge could simply be recast as follows: Let a rational* person be one who is free from confused thinking, prejudice, and similar defects, but who is *not* necessarily omniscient with respect to the a priori. Then Wright's challenge comes to this: it seems plausible that rational* individuals who know the relevant non-moral facts might arrive at incompatible moral views, but this possibility is not something that the moral realist can allow.

But why *can't* the moral realist simply allow that informed and rational (or rational*) individuals might arrive at incompatible moral views? Again, the realist would be committed to denying this possibility if she were committed to the view that rational and informed individuals would inevitably converge on the truth; but as emphasized above, it is unclear why the moral realist should be thought to be committed to the latter view, given that the realist in other domains will be quick to reject the idea that rational inquiry is guaranteed to deliver the truth (even in the

long run). The upshot of the discussion just concluded is that the ostensibly a priori character of morality does not provide a relevant disanalogy.

Let's look elsewhere.

IV. WRIGHT ON REALISM AND REPRESENTATION

According to Wright (1992, 2003), a realist in *any* domain is committed to holding that suitably idealized thinkers would converge on the truth in that domain, in a way that the antirealist is not. Thus, for Wright, the fact that the *moral* realist is committed to an extremely substantial convergence claim is a consequence of a much more general truth about realism. On Wright's account, this commitment flows from the realist's insistence on understanding truth as *accurate representation*. For the realist, the aim of both judgment and sincere assertion is the accurate representation of objective states of affairs. In contrast, the antirealist rejects the idea that truth as accurate representation is the characteristic aim of judgment and assertion; rather, the aim is *superassertibility*, which Wright thinks of as a kind of indefeasible warranted assertibility according to the standards governing the relevant language game.[15]

Here is Wright's characterization of the realist:

> The thought of a realist [about some region of discourse] . . . is that responsibly to practice in that region is to enter into a kind of representational mode of cognitive function, comparable in relevant respects to, say, taking a photograph or making a wax impression of a key. Certain matters stand thus and so independently of us—compare the photographed scene and the contours of the key . . . We engage in a certain process, to wit, we put ourselves at the mercy, so to speak, of the standards of appraisal appropriate to the discourse in question—compare taking the snapshot or impressing the key on the wax—and the result is to leave an imprint in our minds which, in the best case, appropriately matches the independently standing state of affairs. ([1995] 2003, 197)

Let's assume with Wright that the realist is committed to the idea that accurate representation is the aim of assertion and judgment. The crucial question then becomes the following: how does this commitment generate a commitment to any substantial convergence thesis? Here Wright appeals to "a truism connecting representation and convergence":

> . . . that representationally functioning systems, targeted on the same subject matter, can produce divergent output only if working on divergent input or if they function less than perfectly. ([1995] 2003, 146)

In both *Truth and Objectivity* and his later "Truth and Ethics," Wright employs metaphors in order to illustrate and support the relevant line of thought. In what follows, I will argue that the kinds of metaphorical considerations offered by Wright

not only fail to advance his case, but in fact point in the opposite direction when subjected to sufficiently close scrutiny. The one metaphor that appears in both places is the metaphor of the *camera* (2003, 197–98; 1992, 91). For that reason, I will focus my remarks on that metaphor in particular, although I believe that critiques closely analogous to the one that I will offer could be made for any of the other metaphors that Wright employs.

Suppose that both you and I use our cameras to take photographs of the same scene from the same vantage point on some occasion. (Thus, our cameras do not receive "divergent inputs.") If both of our cameras function properly, then the two photographs that result will accurately represent the relevant scene. If, however, the two photographs represent the scene in incompatible ways, then we would know that at least one of the two was a misrepresentation. And in that case, we would be in a position to conclude that at least one of the two cameras had malfunctioned on this occasion.

On the realist's picture, individual believers are like cameras and their judgments are like photographs: the judgments, no less than the photographs, are attempts at accurate representation. Thus, if two individuals with the same background information carefully consider the same moral question, then we would expect them to arrive at the same answer, viz. the *correct* representation. If instead they arrive at incompatible answers, then at least one of the two has misrepresented the actual state of affairs. And in that case, given that the two individuals had the same background information to go on, at least one of the two must have arrived at his answer because of a lapse in rationality, or a failure to conduct his thinking in the way that he should have. (This of course is analogous to the malfunctioning of the camera.) But it seems to follow immediately from this that in a case in which *neither* of the two individuals suffers from any such lapse or absence of background information, they will both arrive at the true view. In Wright's words,

> If we take photographs of the *same* scene which somehow turn out to represent it in incompatible ways, there has to have been some kind of shortcoming in the function of one of the cameras or in the way it was used . . . The price you pay for taking the idea of representation in the serious way the realist wants to take it is that when subjects' representations prove to conflict, then . . . there has to have been something amiss with the way they were arrived at or with their vehicle—. . . the camera, or the thinker. Accordingly, one obligation of the moral realist will be to hold, and therefore to justify holding, that moral disagreements, since they involve a clash of what purport to be substantial representations, have to involve defects of process or materials: at least one of the protagonists has to be guilty of a deficiency in the way he arrives at his view, or to be somehow constitutionally unfit. ([1995] 2003, 198)

But unfortunately for the *moral* realist, it is simply implausible to suppose that a diagnosis along these lines will always be appropriate:

> . . . it is of course evident that moral disagreements can be and frequently are attributable to confused thinking, factual ignorance and

> sheer prejudice. But the [realist's] obligation . . . is to show that such deficiency *has* to be involved in the generation of any such dispute . . . Any student of morality who has come to feel, therefore, that a substantial body of the principles that inform our ordinary moral thought are essentially contestable, and that no rational or cognitive deficiency is needed to sustain the clashes on things like sexual morality, the value of individual freedom, the moral status of animals and the ethics of suicide and mercy-killing, which are freely exemplified within and across cultures, won't give much for the realist's chances. ([1995] 2003, 199)

Before considering how the moral realist might reply, we should take note of an interpretive point that is absolutely crucial to the dialectic between Wright and the realist. (Fortunately, the point in question is one about which Wright himself is completely unequivocal.) In describing the realist's commitments, Wright frequently claims that the realist is committed to thinking that, in cases in which two well-informed individuals arrive at incompatible judgments, at least one of the two individuals must be guilty of a "cognitive deficiency" or (even more commonly) a "cognitive shortcoming." Now, in a sufficiently inclusive sense of "cognitive deficiency" or "cognitive shortcoming," any realist will immediately concede that disagreement about some issue always indicates a deficiency or shortcoming on the part of at least one of the parties to the dispute. For in a context in which the question at issue is whether p is true, there is a clear sense in which *ignorance of whether p* (even if it is perfectly rational ignorance of whether p) itself counts as a cognitive shortcoming.[16] In this sense, even a believer who perfectly proportions her beliefs to her evidence (and is in other ways fully reasonable) has cognitive shortcomings, so long as she falls short of omniscience. (That is, anyone who falls short of literal omniscience possesses cognitive shortcomings that an omniscient God would lack.)

Of course, if mere ignorance about whether p counts as a cognitive shortcoming, then the moral realist will happily concede that individuals without any cognitive shortcomings will inevitably believe all of the truths about morality, and therefore, agree with one another: for that is just to say that beings who are *omniscient* about morality will not hold incompatible moral views. Notice that, on this interpretation, any skeptical argument from the hypothesized lack of convergence is a complete nonstarter. For there is nothing even slightly implausible about the claim that individuals who are omniscient about the domain in question will converge on the truth in that domain; on the contrary, the fact that such individuals will converge on the truth is utterly trivial.

Unsurprisingly then, Wright does not employ "cognitive shortcoming" or "cognitive deficiency" in this inclusive way. Rather, he emphasizes at length that mere ignorance does not count as a cognitive shortcoming or deficiency in the relevant sense.[17] As examples of what *would* count as cognitive shortcomings in the relevant sense, Wright offers "inferential error," "prejudicial assessment of data," "oversight of data," and "dogmatism" (1992, 93). These are all cases of cognitive malfunction, analogous to the way in which a camera might malfunction. But of course, a fully reasonable person would not commit any of these errors, even if she fell well short

of omniscience. The claim that the class of fully reasonable people (who possess relevant background information) would inevitably converge on the truth in some domain is thus a substantive claim, in a way that the claim that there would be convergence among people who are omniscient about that domain is *not* a substantive claim.

Having noted this interpretive point, we are now in a position to consider Wright's claim that the moral realist is committed to some strong convergence thesis in virtue of her commitment to truth as accurate representation.

Let's examine Wright's camera metaphor in more detail. In thinking about the case, it's natural to imagine the cameras being used to take pictures of some relatively nearby person, object, or scene. Perhaps you and I both take a picture of my daughter, who is standing immediately before us. If the resulting photographs represent my daughter in ways that are incompatible—for example, the photograph produced by your camera represents her as having blond hair while the photograph produced by mine represents her hair as brown—then we would know that one of the two photographs misrepresents her. Moreover, in those circumstances, we would treat this as indicative of a malfunction on the part of the camera in question: it's the kind of thing that might lead us to return the camera to the store from which it was purchased, on the grounds that it had failed to do the kind of thing that it was engineered to do.

In *some* cases of the relevant kind then, we clearly do treat the fact that representational devices have produced incompatible representations as a sufficient reason to conclude that one or the other must have malfunctioned or performed defectively in some way. Notice that even if this is *also* true in the moral case, it does not come close to showing that REALISM REQUIRES CONVERGENCE ON THE TRUTH. The analogue in the moral domain would be something like this: with respect to at least *some* moral questions, anyone who was reasonable and knew the relevant nonmoral facts would arrive at the correct answer: therefore, there could not be disagreement among individuals who satisfied the relevant conditions with respect to those questions. This view is extremely plausible. For example, it is extremely plausible that anyone who was fully informed and rational would deny that *we are morally required to occasionally kill randomly*. Therefore, it is extremely plausible that there would be no difference of opinion about this among people who satisfied the relevant conditions. But of course, one might agree with that much while also holding that there are at least some, far more difficult moral questions with respect to which satisfying the relevant conditions does *not* guarantee that one will believe what is true. That is, one might accept that there are at least some moral questions with respect to which all rational and informed individuals would converge on the truth, while denying that REALISM REQUIRES CONVERGENCE ON THE TRUTH.

Consider next a case in which we use our cameras to take photographs of my daughter when she is at a great distance from us. Unsurprisingly, she appears in our pictures as a blurry, indistinct blob. Moreover, perhaps the blob in your picture has a somewhat different shape from the blob in my picture: our photographs thus rep-

resent her as having incompatible properties. Even so, there is no reason at all to think that either of our cameras has malfunctioned. My daughter was simply too far away to be accurately represented by our cameras. Similarly, even if we use our cameras to take photographs of my daughter when she is standing immediately before us, there will be ever so many aspects of her appearance that the resulting photographs will not accurately represent—consider, for example, all of those aspects of the way she appears that are simply below the resolution of either camera. But even though that is true, and even if the way in which your photograph fails to accurately represent some of those aspects is different from the way in which mine fails to accurately represent those aspects, that does not mean that anyone's camera has malfunctioned. Indeed, both cameras might have functioned perfectly well, in a way that would have delighted their manufacturer.

Thus, even when we consider Wright's own example, a representational failure need not be indicative of any *other* failure on the part of the relevant device. In the moral case, the analogue would be this: failure to accurately represent some moral fact does not necessarily indicate any other failure or "cognitive shortcoming" (e.g., a failure of rationality, or nonmoral ignorance) on the part of the person who fails to grasp the relevant moral fact. When we employ our cameras to produce representations of middle-sized dry goods in the immediate vicinity, we expect them to produce accurate representations, and therefore, *convergent* representations. When convergence fails to occur in such cases, we naturally posit malfunction. But even if they are working perfectly, our cameras will of course fail to accurately represent facts that are too distant or that lie beyond their capacities for discrimination. The analogue in the moral case would be this: there are some moral facts that are simply out of range, or beyond our capacities for discrimination, even when we are fully rational and possessed of all relevant nonmoral information.

One might think that it simply makes no sense to think that there might be moral facts that are somehow "out of range" for us.[18] I will attempt to dispel, or at least mitigate, this feeling in the next section. But in the present dialectical context, whether we can ultimately make sense of inaccessible moral facts is not to the point. Recall that, quite apart from anything having to do with *moral* realism in particular, Wright's claim is that *any* realist is committed to a strong convergence claim simply in virtue of her commitment to representationalism. That is, Wright holds that simply by reflecting on the nature of representation, we can see that to take representation seriously (as the realist does) is to incur commitment to a nontrivial convergence-on-the-truth claim: and this holds for realism about the external world, just as much as for moral realism. Against this, I have argued that, when subjected to close inspection, even Wright's own illustrative example points in the opposite direction: commitment to interpreting a domain of discourse in representationalist terms does not carry a commitment to any nontrivial convergence claim. The convergence thesis in the area to which the realist *is* committed is the trivial one: any person who *correctly represents* the relevant subject matter will converge in his or her views with anyone else who correctly represents that same subject matter.

V. UNKNOWABLE MORAL TRUTHS?

Even in the absence of compelling arguments for the claim that a fully rational and informed person would be in a position to know any moral truth, the *feeling* that this is so might remain. That is, one might simply find it grossly implausible that a rational and informed person might nevertheless not be in a position to know some moral fact. For example, Wright frequently compares the suggestion that there might be unknowable moral truths to the suggestion that there might be unknowable *comic* truths (1992, 8–9, 82, 93; [1995] 2003, 192). And for Wright, the domain of the comic is the paradigm of a domain in which it is simply grossly implausible to think that any genuine fact (e.g., about whether a certain joke is funny) could outstrip our ability to know it. In view of this, we should note that there are considerations that suggest that there *are* at least some unknowable moral truths.[19]

Consider, for example, the tendency of borderline cases to create contexts in which even perfectly rational thinkers are not in a position to know the truth.[20] Let's assume that (1) I am morally required to give at least one dollar of my annual income to charity (if I failed to do this, I would be violating a moral obligation that I have) but that (2) I am not morally required to give all of my annual income to charity. Consider then the series of claims "I am morally required to give at least $1 of my income to charity"; "I am morally required to give at least $2 of my income to charity"; . . . "I am morally required to give at least $z of my income to charity" (where z represents my total annual income). If classical logic applies to moral propositions,[21] then there is guaranteed to be some highest number n such that (i) I am morally required to give n dollars to charity, but (ii) I am not morally required to give n+1 dollars to charity. However, even if I'm thinking rationally and am knowledgeable about the nonmoral facts, it doesn't follow that I'm in a position to know what that dollar amount is. (Indeed, one might think that a rational person who is concerned not to violate any of her moral obligations should go safely over the actual value n.)

Williamson (2000) argues powerfully that there are *no* nontrivial conditions that are "luminous," or such that we are always in a position to know whether they obtain (even if we are thinking rationally). He is specifically concerned to argue that even facts about one's current phenomenological states (e.g., whether one feels cold at the current time) can outrun one's capacity to recognize them; he takes these to be the best candidates for genuinely luminous conditions. But as others have noted, Williamson's arguments can be straightforwardly applied to the moral domain. If *I feel cold* is not luminous, then neither is *I am morally required to give more to charity*: there are possible circumstances in which one would not be in a position to know whether one is morally required to give more to charity than one already has, even if one is thinking rationally and is aware of the underlying nonmoral facts.[22]

Does taking seriously the possibility that there might be unknowable moral facts open the door for a more general skepticism about claims to moral knowl-

edge? No more than taking seriously the possibility of unknowable historical or mathematical facts opens the door for a more general skepticism about historical or mathematical knowledge. Even if there are some historical facts that are cognitively inaccessible, that is consistent with our having a great deal of historical knowledge: the fact that we are not in position to know how many hairs were on George Washington's head at the moment of his birth casts no doubt on our knowledge that he lived in the eighteenth century, or that major wars occurred during the twentieth. Similarly, even if there are some moral facts that we are not in a position to know, that in itself would not cast doubt on one's claim to know that it is good to contribute to charity, or that one is not under a standing moral obligation to occasionally kill randomly.

Does taking seriously the possibility that at least some moral facts are unknowable conflict with the appealing idea that there are tight links between moral wrongness and *accountability*? According to this line of thought, countenancing moral facts that are in principle unknowable would open the door for moral wrongs for which no one could be held accountable (even in principle); but this is absurd.[23]

Of course, it is at least somewhat controversial whether there is a significant connection between wrongness and accountability.[24] But let us assume for the sake of argument that there is a significant connection between wrongness and accountability, in order to give the objection its best chance of success.

Suppose that I am not in a position to know the minimum amount of money that I must contribute to charity that is consistent with fulfilling my moral obligation to contribute a sufficient amount. (As argued above, that there *is* such a minimum amount follows immediately from assumptions that many realists will find congenial.) Even if there is simply no way that I could know the relevant moral truth, it does not follow that I cannot be held accountable for failing to give enough to charity. Suppose first that I give nothing to charity, or so little that it *is* knowable that I have not given enough. In that case, I can be held accountable by others for failing to give enough. (The mere fact that there is an unknowable moral truth in the vicinity of my action—even one that is such that, had I known it, I would have acted otherwise—is not enough to show that I cannot be held accountable for my action.) So this is not a case of a wrong for which no one could be held accountable.

Suppose next that I give a certain amount to charity, but that I fall just short of the moral minimum, in such a way that it is in principle unknowable that I have failed to meet it. (Perhaps I desire to give the morally required amount and no more, but because I am necessarily ignorant of what the moral minimum is, I fall just short.) In that case, given that the true moral minimum is unknowable, other people will not be in a position to hold me accountable.[25] But it does not follow that I have not acted wrongly. Imagine that at the end of my life I am confronted by an omniscient god, or an oracle, who informs me of all of the occasions throughout my life in which I should have acted differently. She notes that this

particular year I should have given more money to charity than I did. I might reply as follows: "Perhaps it would have been better for the world if I had given more money than I did. But it cannot be that I acted wrongly, because I simply could not have known the true moral minimum." She should not be impressed with this reply. Among other things, it is consistent with everything that has been specified about the case thus far that there are many amounts that I might have contributed that not only were over the moral minimum, but which were known by me to be over the moral minimum. This story seems perfectly coherent. But now suppose that the same case occurs in a possible world in which (as it happens) there are no omniscient gods or oracles. In that case, I act wrongly, even though there is no one who could hold me accountable. But it is hard to see why it should be thought problematic.

Further, notice that whatever connections hold between wrongness, accountability, and knowability, those connections must be loose enough to accommodate the following possibility: an agent can act wrongly even though the fact that he is acting wrongly is *in practice* something that he is not in a position to know. For example, consider an ancient Hittite lord who inhabits a culture in which a particular kind of slavery is universally regarded as morally unproblematic, even by the slaves themselves.[26] Let's assume that, even in such a culture, it is *in principle* knowable—to the Hittite lord and others—that slavery is wrong. (In principle, they could come to recognize that slavery is wrong in whatever way inhabitants of more enlightened cultures come to recognize this.) Nevertheless, perhaps as things actually stand, the wrongness of slavery is not just *unknown* to the Hittite lord, but is *in practice unknowable* for him: given his limitations and the way in which he is embedded in the world, he is in practice simply incapable of recognizing that slavery is wrong. Even in that case, it is still true that the Hittite lord acts wrongly when he actively participates in and promotes the institution of slavery. In general, an agent can act wrongly even when he acts out of moral ignorance, and the moral truth of which he is ignorant is one that he is not in a position to know, given his limitations. Again, whatever links hold between wrongness, accountability, and knowability, they are at least loose enough to accommodate that possibility.

But then the problem for the objection is this: given that an agent can act wrongly even if he is not in a position to know the relevant moral truths, why should it matter whether those moral truths are unknowable in principle or merely unknowable in practice? (Given that the wrongness of slavery is not something that the Hittite lord is in a position to know about in any case, why should it make a difference how *deep* his inability runs, if the background worry concerns the (im)propriety of anyone's holding him accountable for his behavior?) On the face of it, there does not seem to be any reason to think that unknowability in principle is incompatible with wrongful action that is not *also* a reason for thinking that unknowability in practice is incompatible with wrongful action. But as we have noted, the fact that an agent acted in ignorance of a moral truth that was in practice unknowable for him is compatible with his having acted wrongly. Therefore, an agent's acting in igno-

rance of a moral truth that was in principle unknowable for him is compatible with his having acted wrongly as well.[27]

So far we have examined a number of different ways of motivating REALISM REQUIRES CONVERGENCE ON THE TRUTH and found them wanting. The next section turns to Michael Smith's argument from the objectivity of moral judgment.

VI. SMITH ON THE OBJECTIVITY OF MORAL JUDGMENT

For Smith, one of the most important features of morality is that moral judgments at least purport to be *objective*. In the opening pages of *The Moral Problem*, he offers a characterization of what this amounts to:

> We may summarize this . . . feature of morality in the following terms: we seem to think that moral questions have correct answers; that the correct answers are made correct by objective moral facts; that moral facts are wholly determined by circumstances; and that, by engaging in moral conversation and argument, we can discover what these objective moral facts determined by the circumstances are. The term "objective" here simply signifies the possibility of a convergence in moral views of [a certain kind]. Let's call this the "objectivity of moral judgment." (1994, 6)[28]

Significantly, Smith holds that the kind of possible convergence in our moral views that is presupposed by the objectivity of moral judgment is convergence on the truth via rational inquiry:

> A careful mustering and assessment of the reasons for and against our particular moral opinions about [moral] dilemmas and issues is . . . the best way to discover what the moral facts really are. If we are open-minded and thinking clearly then such an argument should result in a convergence in moral opinion, a convergence upon the truth. (5–6)

When taken as a rough-and-ready generalization about the efficacy of open-mindedness and clear thinking, this last remark about convergence seems unobjectionable. (Indeed, it seems to have the platitudinous status that Smith claims for it.) However, Smith holds a significantly stronger thesis about the connection between convergence among rational thinkers and the moral facts. It is not simply that fully rational thinkers would *often* converge on the truth in the moral domain, or that they are much more likely to do so in virtue of their rationality. Rather, Smith's view is that *whatever moral facts there are*, are facts on which all fully rational thinkers would converge.[29] That is, if fully rational thinkers would not all agree that some moral claim F is true, then it follows immediately that F is no moral fact at all.

I think that Smith is right in thinking that ordinary moral judgments purport to be objective and that this is a fact of deep significance for metaethics. But I think that it is a mistake to think that the objectivity of moral judgment entails any

significant (i.e., nontrivial) convergence claim. I take it that one thing that is often meant when it is claimed that moral judgments purport to be *objective* is that in this respect such judgments do not differ from the purely descriptive, nonnormative claims that are routinely made in the natural and social sciences. In any case, even the most thorough-going moral realist need not claim that moral judgments aspire to a kind of objectivity that goes beyond the objectivity of paradigmatic judgments in physics, or biology, or history. Thus, one strategy for showing that the "objectivity of moral judgment" does not entail or presuppose any significant convergence thesis is to show that the objectivity of paradigmatic scientific judgments does not presuppose or entail any significant convergence thesis. This is the strategy that I will pursue in what follows.

Consider paradigmatic historical judgments such as the following:

(1)   The Battle of Hastings occurred in the year 1066.

(2)   George Washington was the first president of the United States.

(3)   Over a million men were killed or wounded in the Battle of the Somme.

These paradigmatic historical judgments purport to be objective. What does this amount to? Recall the first two features that Smith lists in his characterization of "the objectivity of moral judgment":

> . . . we seem to think that moral questions have correct answers; [and] that the correct answers are made correct by objective moral facts . . .

Suppose that a philosopher concerned to explicate "the objectivity of historical judgment" followed Smith's lead:

> . . . we seem to think that historical questions have correct answers, and that the correct answers are made correct by objective historical facts

but then *stopped* before saying anything more. My claim is that this would be a *sufficient* explication of the "objectivity of historical judgment" (i.e., the way in which paradigmatic historical judgments at least purport to be objective), and that it does not entail or presuppose anything interesting about convergence. Indeed, suppose that the philosopher proceeded to give the following speech:

> Not only do I think that historical judgments *purport* to be objective, I think that they *are* objective. What I mean by that is simply the following. First, questions like "When did the Battle of Hastings occur?" or "Who was the first President of the United States?" have correct answers. Second, those correct answers are made correct by objective historical facts. In some cases, we have access to the objective historical facts, and so know what the correct answers are. But in other cases, we don't. Moreover, in many cases of the latter kind, our ignorance would be shared by even perfectly rational historical inquirers. Consider, for example the following historical question: "Did George Washington

have an even number of hairs on his head at the time of his birth?" I believe that there is a correct answer to this question, and that this answer is made by an objective historical fact. Of course, I don't think that I know what that correct answer is, and I believe that my ignorance would be shared even if I were perfectly rational, and had access to every piece of evidence that actually exists that could conceivably be thought relevant to answering the question. Moreover, I think that there are countless other historical facts—some possessing much greater historical interest—of which the same is true. Therefore, I believe that there are countless truths of history on which fully rational historical inquirers would fail to converge. But that is perfectly consistent with my belief in the objectivity of history. Indeed, I regard the idea that there are no such historical facts as a characteristic error of the *antirealist* about history. I have a low opinion of antirealist views about history, but I won't bore you with my reasons for that attitude today. For current purposes, the important point is simply that my own commitment to the objectivity of history and the existence of objective historical facts in no way commits me to any nontrivial convergence claim to the effect that these historical facts would be known to fully rational historical inquirers.

But of course, a moral realist might give a perfectly parallel speech in explicating her commitment to the objectivity of morality.

As before, it is not a relevant difference that historical judgments are uncontroversially empirical judgments while moral judgments are not. After all, a parallel speech could also be given by a mathematical realist, a speech in which lines like the following occur:

> The mathematical question, "Is Goldbach's Conjecture true?" has a correct answer, a correct answer that is made true by an objective mathematical fact, even if there is no route by which even perfectly rational inquirers could ever come to know it.

In short, a mathematical realist might be firmly committed to the actual (and not just purported) objectivity of mathematics without being committed to any nontrivial convergence-on-the-truth thesis. But if morality can be objective in the same way that mathematics is, that is surely objectivity enough: we should not build in additional requirements for moral objectivity, beyond those that apply to mathematics and mathematical discourse.

Why does Smith build in the requirement of convergence in his explication of objectivity in the moral domain? One possibility is this. Notice that a philosopher who delivers a speech of the kind given above—whether the speech concerns history, mathematics, or morality—takes for granted that there is a robust distinction between *rational believing* and *believing what is true* in the domain in question. But it is not clear that Smith accepts this distinction. In several places, he suggests that in order to qualify as fully rational, an individual must not harbor any false beliefs, or lack any true beliefs that are relevant to the issue at hand. Consider, for example, the following characterization of the "fully rational agent":

> My suggestion is that to be fully rational an agent must not be suffering from the effects of any physical or emotional disturbance, she must have no false beliefs, she must have all relevant true beliefs, and she must have a systematically justifiable set of desires ([1997] 2004, 263).[30]

Copp (1997, 44) criticizes Smith on the grounds that this account *connects rational believing* and *believing what is true* too closely. In response, Smith suggests that "the term 'rationality' is almost entirely a philosopher's term of art," and that, in any case, he doubts whether anything hangs on the way in which he uses the term ([1997] 2004, 265–266).

Suppose that we simply agree to use "fully rational" in the way that Smith suggests, so that someone who is fully rational will have no false beliefs and all relevant true beliefs about the subject matter in question. Once that stipulation is in place, any moral realist should immediately agree that

(1)   If A and B disagree about the morality of φ-ing in circumstances C, then at least one of A and B is not fully rational.

For this claim is now being understood in a way that adds nothing to the claim that there is a fact of the matter about whether φ-ing is right in circumstances C. That is, given that there is an objective fact of the matter about whether φ-ing in circumstances C is right, and having a false belief about what that objective fact of the matter is entails that one is not "fully rational," then it follows trivially that there will be no disagreement about what that fact of the matter is among the fully rational. Moreover, given that anyone who is "fully rational" will not lack any relevant true beliefs, it follows immediately that any fully rational person will not only lack the false belief about whether φ-ing is right, but will also possess the true belief about whether it is. Thus, any moral realist should agree that

> Any fully rational person will believe the truth about whether φ-ing in circumstances C is right.

Given that "φ-ing in circumstances C" is being used as an arbitrary stand-in for any action that might be assessed with respect to its morality, any moral realist should agree with the more general claim that

> Any fully rational person will believe the truths about morality.

But again, the moral realist should see *this* convergence claim as utterly trivial, in the following sense: it adds nothing to that to which he was already committed by his belief in objective moral facts.

One way of appreciating this triviality is to note the following. A person who had no false beliefs and all relevant true beliefs about some topic would not differ in her beliefs from someone who was *omniscient* about that topic.[31] But as we noted in section 4, it is utterly trivial that anyone who was omniscient in some domain would believe all of the truths in that domain, and so not disagree with anyone else who was omniscient.

Let us summarize the conclusions of this section. First, as long as one distinguishes sharply between rational believing and believing what is true, there is no route from the *objectivity of morality* to the claim that rational thinkers will inevitably converge on the moral facts. In particular, a moral realist can consistently uphold the objectivity of morality, while denying that there is any guarantee (or even any reason to think that) any moral fact is something on which all rational and informed thinkers will ultimately converge. On the other hand, if one uses "fully rational" in such a way that being a fully rational believer entails not believing what is false and believing what is true, then the moral realist *is* committed to the claim that fully rational believers will converge on the moral facts. But this commitment is trivial, in the sense that it adds nothing to a claim not involving convergence to which the realist was already committed, viz. that there are objective moral facts.

VII. INTERPERSONAL CONVERGENCE WITHOUT CONVERGENCE ON THE TRUTH?

One way in which REALISM REQUIRES CONVERGENCE could be true is if the stronger claim that REALISM REQUIRES CONVERGENCE ON THE TRUTH is. But REALISM REQUIRES CONVERGENCE might be true even if the stronger claim is not. Suppose that whenever a rational and informed person fails to believe the truth about a given moral issue, any other rational and informed person would share whatever position he does take on that issue. In that case, rational people who were fully informed about the nonmoral facts would not hold incompatible moral views, even if there were some moral issues with respect to which they failed to believe the truth.

In fact, the kinds of considerations discussed in previous sections might very well encourage the idea that this is the situation. For example, I have suggested that just as the realist about the past will insist that there is a potential gap between historical facts and what even idealized historical inquirers are in a position to know about the past, so too the moral realist should insist that there is a potential gap between the moral facts and what even perfectly rational and informed individuals are in a position to know about morality. But significantly, the kinds of stock examples that are offered to make realism about the past seem compelling do *not* motivate the idea that rational individuals could differ in what they believe about those historical claims. For example, if there is currently no evidence that bears on the question of whether George Washington had an odd or an even number of hairs on his head at the moment of his birth, then rational people who consider the question will respond in the same way: by suspending judgment. The same point holds for ostensibly a priori domains such as mathematics. Plausibly, agnosticism is the rational stance toward Goldbach's Conjecture in the absence of either a proof or a counterexample; on the realist assumption that Goldbach's Conjecture is either

true or false, rational individuals will fail to believe the truth about Goldbach's Conjecture. But because they adopt the same stance, they will not disagree about whether it is true or false. More generally, because any case of *underdetermination* will rationally require an attitude of agnosticism, any case of underdetermination will be one in which rational individuals neither believe the truth nor disagree.

The same holds for other clear cases in which rational believers fail to believe the truth, e.g., cases in which the evidence is misleading. If all of the remaining evidence misleadingly suggests that some historical truth is false, then rational thinkers will fail to believe the truth in virtue of responding rationally to their evidence. But they will still converge in their views, for they will all believe what is false. In general, the kinds of considerations that are used to motivate the thought that truth is not epistemically constrained in domains such as the historical past or mathematics do not motivate the view that rational and informed inquirers might arrive at incompatible views. On the contrary, the cases in question seem to suggest that rational individuals would not differ in their views even in those cases in which they fall short of the truth.

Similarly, even if my fully rational and informed self would not be in a position to know the moral truth that "SM is required to give $n to charity" because of its status as a borderline case, it doesn't follow that there would be any disagreement about it among informed and rational people. (Perhaps we should all be agnostic about its truth.)

In short, many of the best reasons for skepticism about REALISM REQUIRES CONVERGENCE ON THE TRUTH are *not* also reasons for skepticism about REALISM REQUIRES CONVERGENCE. Thus, although philosophers who accept REALISM REQUIRES CONVERGENCE generally do so *because* they accept REALISM REQUIRES CONVERGENCE ON THE TRUTH, I believe that the best case for the former does not employ the latter as a premise. That having been said, there are reasons for skepticism about REALISM REQUIRES CONVERGENCE as well. In the remainder of the paper, I'll sketch one of these reasons.

If each of us arrived at his or her moral views in the most rational way possible would we arrive at the same views? One obstacle to answering this question is that there is no generally agreed-upon account of how it is rational to arrive at one's moral views. However, many moral philosophers, including both Smith and Rawls, hold that the rational way to make up one's mind about moral questions is via the *method of reflective equilibrium* (Goodman 1953; Rawls 1971).[32] Indeed, prominent moral philosophers sometimes suggest that when it comes to moral inquiry, the method of reflective equilibrium is, in effect, the only game in town.[33] Given this, it seems that the following question is worth asking:

> On the assumption that the method of reflective equilibrium is a reasonable way to make up one's mind about moral questions, could different people, all of whom are aware of the nonmoral facts, rationally arrive at incompatible moral views?

In the remainder of this section, I will argue that the answer to this question is "yes": we should not expect that different thinkers who follow the method of reflective equilibrium—even impeccably—will arrive at the same moral views. But if that is correct, then realists who think that reflective equilibrium is at least *a* reasonable way to arrive at one's moral views should not endorse REALISM REQUIRES CONVERGENCE.[34]

Consider a representative characterization of the method by its most influential proponent, Rawls:

> People have considered judgments [about morality] at all levels of generality, from those about particular situations and institutions up through broad standards and first principles to formal and abstract conditions on moral conceptions. One tries . . . to fit these various convictions into one coherent scheme, each considered judgment whatever its level having a certain initial credibility. By dropping and revising some, by reformulating and expanding others, one supposes that a systematic organization can be found. Although in order to get started various judgments are viewed as firm enough to be taken provisionally as fixed points, there are no judgments of any level of generality that are in principle immune to revision. (1974, 289)

By proceeding in this way, one attempts to bring one's moral convictions into a state of reflective equilibrium. On Rawls's account, the state that we should pursue is one of *wide* (as opposed to "narrow") reflective equilibrium. The pursuit of wide reflective equilibrium is the pursuit of a comprehensive moral view that "would survive the rational consideration of all feasible moral conceptions and all reasonable arguments for them" (1974, 289).[35] Of course, Rawls acknowledges that it is not realistic that we will actually consider all such conceptions and arguments.[36] Rather, the state of wide reflective equilibrium constitutes an ideal: it is the hypothetical end point of properly conducted moral inquiry, if such inquiry were pursued without limit.

Rawls himself was much concerned with questions about whether the method of reflective equilibrium would lead to a convergence among those who employed it. In *A Theory of Justice*, he raised, but did not pursue, the following issues:

> This explanation of reflective equilibrium suggests straightaway a number of further questions. For example, does a reflective equilibrium (in the sense of the philosophical ideal) exist? If so, is it unique? Even if it is unique, can it be reached? Perhaps the judgments from which we begin, or the course of reflection itself (or both), affect the resting point, if any, that we eventually achieve. (50)

Consider the issue of whether there is a unique reflective equilibrium. Presumably, there are at least two questions here:

(1) The intrapersonal question: for any particular person, is there some unique reflective equilibrium that she would arrive at if she employed the method impeccably?

(2) The interpersonal question: would different individuals, each of whom employed the method impeccably, converge on a unique reflective equilibrium?

Consider first question (1). Given that one's considered moral judgments are currently not in equilibrium, is there any reason to suppose that there is some rationally optimal way for one to resolve those conflicts that exist? On the face of it, it seems that there might be multiple ways of pursuing coherence, ways that would result in at least somewhat (and perhaps even radically) different sets of judgments.[37] Of course, what is relevant here is *wide* reflective equilibrium. Perhaps if one were presented with "all feasible moral conceptions and all reasonable arguments for them," one would feel rationally compelled to resolve those conflicts in exactly one way, and be driven to some specific equilibrium. Although this is far from obvious, let's simply assume that that's what would happen; more generally, let's assume for the sake of argument that the answer to question (1) is "Yes."

Still, it doesn't follow that for different individuals there is some unique reflective equilibrium. In general, that (1) receives an affirmative answer is a necessary but insufficient condition for (2)'s receiving an affirmative answer. If the answer to (1) is affirmative, then, for any particular set of initial considered judgments that a person might hold, there is some unique reflective equilibrium that would be reached by impeccably applying the revision procedure to that set. Even if that's true, it of course doesn't follow that impeccably applying the procedure to a different set of initial starting points would lead to the same state. Indeed, on the face of it, this seems extremely unlikely. Perhaps the following is among one's considered moral judgments:

> Even if a doctor could save the lives of two people dying for want of some vital organ by forcibly overpowering and harvesting the organs of some innocent and unwilling bystander, it is morally impermissible for her to do so.

If so, then in all likelihood, one also holds other considered judgments with which this judgment coheres. As we have seen, in Rawls's scheme, one is supposed to treat each of these judgments as having "a certain level of initial credibility." Someone with act utilitarian sympathies might have, among his considered judgments, the judgment that in the envisaged scenario the doctor is not only permitted but *morally required* to harvest the organs of the bystander; no doubt, that judgment coheres well with other things that he believes. Given these radical differences, it seems extremely unlikely that the best way for each individual to achieve coherence among his or her *own* judgments will lead them to converge on the same set of views. That is, it seems plausible that:

> Different individuals might impeccably employ the method of reflective equilibrium and end up with substantially different moral views.

Elsewhere, Thomas Kelly and I have argued that there are other reasons to believe this claim, in addition to the simple plausibility considerations offered here.[38]

One might object that the line of thought offered here presupposes an overly individualistic conception of the method of reflective equilibrium. Perhaps if there is any significant disagreement about a given moral judgment (e.g., that it is impermissible to sacrifice the two innocent people in the organ-harvesting case), then that judgment is *ipso facto* disqualified as a permissible starting point. That is, perhaps it is a necessary condition of a judgment's being a "considered judgment" in the relevant sense that there is *general agreement* that it is true.[39] Indeed, if one imposed a sufficiently strong social requirement on considered judgments, then one might be led to claim that there is in fact a *uniquely permissible* starting point for the method of reflective equilibrium.[40]

However, there are good reasons to reject the requirement that permissible starting points must be generally accepted. Consider, for example, what we might call *the knowledge platitude*:

> The Knowledge Platitude: If one knows something that is relevant to a question that one is trying to answer, one should take that information into account in arriving at a view (Kelly 2008, 65; Cf. Williamson 2000, ch. 9, and 2007, ch. 7).

The knowledge platitude is inconsistent with strong social requirements on permissible starting points. For example, if one genuinely knows something that is not generally accepted (suppose that many people do accept the claim in question, although not everyone does), then a norm to the effect that one should set aside any judgment that is not generally accepted would have one set aside this piece of knowledge. If that piece of knowledge is relevant to the issue about which one is thinking, then one will have violated the knowledge platitude. For example, perhaps before beginning the process of seeking reflective equilibrium among one's moral views, one genuinely knows that φ-*ing is wrong*. (Perhaps one learned that φ-ing is wrong from one's parents, who knew it themselves.) But if the wrongness of φ-ing is not generally accepted, then a norm that would have one set aside any such judgment would have one ignore one's knowledge that φ-ing is wrong. This would involve violating the knowledge platitude, because one's knowledge that φ-ing is wrong is clearly relevant if one is attempting to determine (e.g.) which general principles about wrongness one should accept. Since the knowledge platitude is true, we should not accept putative norms that would lead to its violation. Therefore, the best version of the method of reflective equilibrium will not make it a necessary condition on permissible starting points that they are generally accepted.

The method of reflective equilibrium is typically offered as an account of how moral inquiry should proceed, as opposed to an account of the conditions under which particular moral beliefs are justified or reasonable to hold. However, given that proponents of the method take the achievement of wide reflective equilibrium to be the goal of properly conducted moral inquiry, it's natural to treat the following as a *sufficient* condition:

> If an individual achieves a state of wide reflective equilibrium by impeccably applying the method, then the moral views that she holds in that state are reasonable ones for her to hold.[41]

From this claim, and the claim that different people might arrive at incompatible moral views by employing the method, it follows that different people could reasonably arrive at incompatible moral views. So proponents of reflective equilibrium should reject:

> **CONVERGENCE**: Rational individuals who were fully informed about the nonmoral facts would not hold incompatible moral views.

Of course, that leaves open the possibility that a proponent of reflective equilibrium who denies CONVERGENCE is thereby committed to denying moral realism. As noted above, Rawls himself seemed to think that the very existence of "objective moral truths" presupposes that there is a unique wide reflective equilibrium, or at least, that any differences between moral views affirmed in wide reflective equilibrium would be relatively marginal (1974, 290; 1999, 301). Indeed, the fact that he repeatedly and quite self-consciously eschewed talk of truth in the moral domain seems to have been at least in part due to this view, combined with increasing skepticism about whether diverse individuals competently pursuing reflective equilibrium would ultimately converge in their substantive moral views. I think that Rawls was right to be skeptical about the existence of a unique wide reflective equilibrium across different individuals but wrong to assume that moral realism (or "objective moral truths") requires this. For in general, it is perfectly possible to combine realism about a given domain with a commitment to the method of reflective equilibrium as the correct procedure for investigating that domain.

Consider, for example, David Lewis on *philosophy*:

> Our "intuitions" are simply opinions: our philosophical theories are the same. Some are commonsensical, some are sophisticated; some are particular; some general; some are more firmly held, some less. But they are all opinions, and a reasonable goal for a philosopher is to bring them into equilibrium. Our common task it to find out what equilibria there are that can withstand examination, but it remains for each of us to come to rest at one or another of them . . .
>
> Once the menu of well worked-out theories is before us, philosophy is a matter of opinion. Is that to say that there is no truth to be had? Or that the truth is of our own making, and different ones of us can make it differently? Not at all! If you say flatly that there is no god, and I say that there are countless gods but none of them are our worldmates, then it may be that neither of us is making any mistake of method. We may each be bringing our opinions to equilibrium in the most careful possible way, taking account of all the arguments, distinctions, and counterexamples. But one of us, at least, is making a mistake of fact. Which one is wrong depends on what there is. (1983, x–xi)

Plausibly, what Lewis thought is true of philosophy in general holds also for the moral domain: even if different people, each of whom is pursuing equilibrium

among her opinions in a rationally impeccably manner, ultimately settle on different equilibria, this is no reason to doubt that that there is an objective matter of fact that divides them. So if the method of reflective equilibrium is a reasonable way to arrive at one's moral views, we should reject REALISM REQUIRES CONVERGENCE.

Of course, a philosopher who defends REALISM REQUIRES CONVERGENCE might argue that the method of reflective equilibrium is *not* a reasonable way to arrive at one's moral views, and that the correct account of how to do so precludes moral disagreement among individuals who satisfy the relevant conditions. In the absence of such a demonstration, however, we lack good reasons to think that REALISM REQUIRES CONVERGENCE is true.[42]

NOTES

1. Note that even if we are confident that a particular moral dispute is sustained by nonmoral ignorance, this should not make us overly sanguine about its eventual resolution. Generally speaking, the fact that a moral dispute is contingent on nonmoral ignorance does not mean that the dispute is unlikely to persist, or even that it is less likely to persist than moral disagreements that are *not* contingent on nonmoral ignorance. Perhaps some moral disagreements depend on nonmoral ignorance, the elimination of which would require social scientists to conduct large-scale controlled experiments that are simply not practically feasible. Indeed, even if the persistence of a given moral dispute is contingent on the persistence of underlying nonmoral ignorance, this does not guarantee that the moral dispute is even *in principle* resolvable, for the underlying nonmoral ignorance might itself be in principle impossible to eliminate.

2. A particularly optimistic assessment is Boyd (1988, 213), who claims: "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." Other relatively optimistic assessments include Brink (1989, 197–209) and Scanlon (1999, 354–51). Contrast Enoch (2009, 24), Harman (1999), and Mackie (1977, 37–38).

3. The fact that diverse individuals cannot be expected to converge on the same substantive moral views in wide reflective equilibrium, and the political consequences of this, is one of the driving themes of Rawls (1993).

4. As he puts it:

   > . . . we must remember that where entrenched disagreements currently seem utterly intractable we can often explain why this is the case in ways that make them seem less threatening to the idea of a convergence in the opinions of fully rational creatures. For example, one or the other parties to the disagreement all too often forms their moral beliefs in response to the directives of a religious authority rather than as the result of the exercise of their own free thought in concert with their fellows. But beliefs formed exclusively in this way have dubious rational credentials . . . The fact that disagreement persists for this sort of reason thus casts no doubt on the possibility of an agreement were we to engage in free and rational debate. (1994, 188–89)

   Cf. Brink (1989, 205–6).

5. See the helpful discussion of such mechanisms in Enoch (2009, 24–28) and the many references provided there.

6. Smith himself seems to admit as much in various places. On this point, see especially his reply to Enoch (2007) in Smith (2007).

7. Compare the position of the libertarian about free will, who is committed to thinking that the physics of our world is indeterministic. If agnosticism was currently the most reasonable stance about whether the physics of our world is deterministic or indeterministic, then libertarianism would not be a reasonable stance; one should at most be agnostic about its truth.

8. In fact, I'm unaware of any philosopher who accepts REALISM REQUIRES CONVERGENCE but who does not accept the logically stronger claim.

9. As will become clear, my attention to this last question is not simply a matter of wanting to close an unattractive and unoccupied place in logical space for the sake of completeness. On the contrary, for reasons that will emerge, I believe that the best defense of REALISM REQUIRES CONVERGENCE proceeds along the following lines: all of the circumstances in which rational and informed individuals fail to converge on the truth about morality are also circumstances in which they do not end up with different moral views.

10. Two famous examples of the latter: C. S. Peirce's (1940) view that truth is the opinion on which scientists would converge in the hypothetical limit of scientific inquiry and Hilary Putnam's (1981) "internal realism," according to which truth is identical with rational acceptability in ideal epistemic circumstances.

11. For example, Thomas Nagel, in the course of defending "normative realism," claims that "The truth here could not be radically inaccessible in the way that the truth about the physical world might be" (1986, 186). Compare p. 139:

> The connection between objectivity and truth is . . . closer in ethics than it is in science. I do not believe that the truth about how we should live could extend radically beyond any capacity we might have to discover it (apart from its dependence on nonevaluative facts we might be unable to discover) . . . Ethical thought is the process of bringing objectivity to bear on the will, and the only thing that I can think of to say about ethical truth in general is that it must be a possible result of this process, correctly carried out.

12. According to Goldbach's Conjecture, every even number greater than the number 2 is the sum of two prime numbers.

13. To be fair, that usage certainly has its precedents. For example, there is a long (and continuing) tradition according to which any fully rational individual is at least logically omniscient.

14. *Objection*: the failure of a contemporary mathematician to believe the most ingenious results of twenty-third-century mathematics *does* have something important in common with cases in which (e.g.) one irrationally fails to draw an obvious and important conclusion from one's current knowledge. In both cases, there is a legitimate route to a conclusion of interest that the person fails to follow. So if contemporary mathematicians were *perfectly* rational, they would already believe the relevant theorems. Therefore, perfect rationality *does* entail omniscience with respect to mathematics, and more generally, any domain that is a priori. (Thanks to Elizabeth Harman for pressing a point along these lines.)

Reply: Even granting the truth of the premise, the ultimate conclusion does not follow. Suppose that Goldbach's Conjecture is both (i) true and (ii) not provable, even in principle. (Again, this is a possibility for which any mathematical realist will allow.) In that case, there would be no *route* by which the truth of Goldbach's Conjecture could be discovered, even in principle. Why then would perfect rationality entail knowledge of Goldbach's Conjecture, given that perfect rationality does not entail knowledge of (e.g.) historical facts about the distant past for which there is simply no evidence, or any possible route by which such historical facts might be discovered? Someone might insist that perfect rationality entails knowledge of Goldbach's Conjecture (but not unknowable historical facts) because Goldbach's Conjecture is a mathematical truth, and many *other* mathematical truths are knowable on the basis of reason alone (but no historical facts are knowable on the basis of reason alone). However, this move seems completely unmotivated, and would make the concept of rationality a bizarrely gerrymandered one.

An even more radical suggestion (and one which is not vulnerable to the 'gerrymandering' charge) is that perfect rationality entails *omniscience across the board* (e.g., full knowledge of the historical as well as the mathematical facts). Something like this is perhaps Smith's (1994) view; on this suggestion, see section 6 below.