Moral Disagreement among Philosophers

Ralph Wedgwood

1. An argument from disagreement

Many different arguments have been developed by philosophers, over the years, for the conclusion that the existence of widespread moral disagreement creates serious problems for non-sceptical moral realism. In this talk, I shall respond to what seems to me to be one of the most challenging arguments of this kind.1

For the purposes of this essay, “moral realism” will be taken to be the view that moral beliefs have non-relativistic truth-values. According to this view, the content of every moral belief is a moral proposition, and every such moral proposition is either true or false – that is, true or false simpliciter, without relativization to anything else. A version of moral realism counts as “non-sceptical” if and only if it does not make it implausible to claim that a reasonably large number of ordinary thinkers know a reasonably large number of moral truths. (Clearly, it is vague how many thinkers and how many moral truths count as a “reasonably large number”. So this characterization of what it is for a version of moral realism to be “non-sceptical” is vague. However, a vague characterization of this kind will be sufficient for our purposes.)

Let us interpret “disagreement” in the following way. We shall say that one thinker “disagrees” with a second if and only if the first thinker believes a proposition $p$, and the ________________

1 For an illuminating survey of many of these arguments, see Enoch (2011: ch. 8). I have elsewhere commented on a rather different argument from moral disagreement; see Wedgwood (2010).
second thinker believes a second proposition \( q \) that is logically incompatible with \( p \). (We may interpret the notion of “believing a proposition \( p \)” in a broad and capacious way, to involve having a level of confidence in \( p \) that is at least higher than one’s level of confidence in \( p \)’s negation, \( \neg p \).) Given this understanding of disagreement, non-sceptical moral realism implies that whenever two thinkers disagree about a moral issue, at least one of those two thinkers believes something \textit{false}, and so holds an \textit{incorrect} or \textit{mistaken} belief.

Clearly, any theorists who claim that moral disagreement creates serious problems for non-sceptical moral realism must somehow argue for the following thesis:

\[
\text{If non-sceptical moral realism were true, such disagreements would not arise – or at least they would not arise in the precise way in which they actually do.}
\]

It is not clear offhand exactly how these theorists will argue for this thesis. It is not as if there is any plausibility in the claim that the truth must necessarily reveal itself to all thinkers’ minds, or that thinkers cannot differ from each other quite significantly in the beliefs that they hold.\(^2\) In this chapter, I shall try to reconstruct the most persuasive argument that I can think of for this crucial thesis.

The theorists who argue that disagreement creates problems for non-sceptical moral realism often claim that such realist views cannot be reconciled with what seems to be the correct \textit{explanation} of this sort of disagreement. In effect, they argue that if non-sceptical moral realism were true, it would impose certain \textit{conditions} on the explanation of such disagreements.

\(^2\) As Enoch (2011: 190) says, “the point can quite plausibly be made that given our cognitive shortcomings agreement rather than disagreement is what calls for explanation, that quite generally disagreement is what you should expect, and agreement the surprising exception that cannot be accepted as brute.”
disagreements – but it is clear that no correct explanation of these disagreements meets these conditions.

To fix ideas, let us focus on cases of disagreement in which the proposition $q$ believed by the second thinker is not just logically incompatible with $p$ (the proposition that is believed by the first thinker), but is simply the negation of $p$ – that is, $\neg p$. In these cases, classical logic guarantees that one of the two thinkers believes something true while the other thinker believes something false. Just to give them labels, let us call the thinker who believes the true proposition “Right-thinker”, and the thinker who believes the false proposition “Wrong-thinker”. The question that the proponent of this argument focuses on is the following: What explains why Right-thinker believes the truth about this issue, while Wrong-thinker believes something false?

One answer might be that it is simply a fluke that Right-thinker believes the true proposition while Wrong-thinker believes a false proposition. In that case, however, neither side in this disagreement can possibly know the proposition that they believe. Obviously, Wrong-thinker cannot know the proposition that he believes, because that proposition is false. But likewise, it seems that Right-thinker also cannot know the proposition that she believes if it a sheer fluke that she arrived at the truth.

So, if either of these two thinkers is to know the truth about this moral issue, it must not be a mere fluke that Right-thinker believes the truth. There must be an explanation of why Right-thinker believes the truth that appeals to some sort of cognitive capacity or intellectual virtue that Right-thinker deploys in holding this belief, which in Right-thinker’s circumstances is sufficient to explain why she believes the truth – while Wrong-thinker either fails to deploy that cognitive capacity or virtue, or else deploys the capacity or virtue in circumstances that
are sufficiently unfavourable to explain why he arrives at a false belief. In short, if either of these two thinkers is to have knowledge, there must be an asymmetry between these two thinkers: one of them (Right-thinker) must have suitable cognitive capacities or virtues and be deploying them in a relevantly favourable environment, while the other thinker (Wrong-thinker) must be either failing to deploy suitable capacities and virtues, or else deploying them in a relevantly unfavourable environment.

Theorists who wish to use this sort of argument to raise problems for non-sceptical moral realism typically claim that it is not plausible to postulate any such asymmetry between the two thinkers. If there is no asymmetry of this sort, then either (a) there is no non-relative truth that the two sides are disagreeing about, or else (b) neither side of the disagreement has knowledge. Either way, this diagnosis of the case seems to support the conclusion that non-sceptical moral realism faces some significant problems.

One potential problem with this argument is that it rests on the claim that it is not plausible to identify any asymmetry between the two sides of the disagreement. In fact, however, in most cases of moral disagreement it is easy to identify such an asymmetry between the two sides. For example, Aristotle (Politics, I.iii-vii) argues for the view that certain forms of slavery are perfectly just. On this point, Aristotle disagrees with most contemporary Western thinkers. But it is clear that Aristotle’s view on this point rests on certain non-moral beliefs. Specifically, his view rests on his non-moral belief that non-Greeks are mentally inferior to Greeks to such a degree that it is actually in the interests of non-Greeks to live as the slaves (that is, the “human tools”) of Greeks. As it happens, these non-moral beliefs are false, and

3 A necessary condition on knowledge of this kind would be accepted by many different epistemologists, although the particular formulation that I give here is inspired by Sosa (2011).
have been decisively refuted by empirical investigations that are accessible to all educated people today. To that extent, even if Aristotle’s capacities for specifically moral thinking were in no way inferior to ours, he was deploying those capacities in significantly less favourable circumstances than we are. So there is no difficulty in identifying an asymmetry between Aristotle and us when it comes to our views about the justice of slavery.

Moreover, we should also not erect artificial barriers to the identification of an asymmetry of this sort. In particular, a correct identification of this sort of asymmetry between the explanations of the beliefs on the two sides of the disagreement does not have to be in any way “neutral” between these two sides; it certainly does not have to be capable of persuading either side of this disagreement to abandon their view. On the contrary, the explanation of why it is that one side of the disagreement is mistaken will be at least as controversial as the claim that that side of the disagreement is mistaken. This should not seem surprising. In general, if it is controversial whether or not a certain fact obtains, the correct explanation of why that fact obtains will also be at least equally controversial. The non-sceptical moral realist does not have to give a non-controversial identification of this sort of asymmetry between the explanations of the beliefs on the two sides of the disagreement. It is a plausible condition on knowledge that there should be a correct explanation, of a certain sort, of why the relevant thinker believes the truth on the relevant question; it is not a plausible condition on knowledge that this explanation should be one that everyone will accept.

So the challenge for the proponent of this argument against non-sceptical forms of moral realism is to identify cases of disagreement that cannot be dealt with as easily as our disagreement with Aristotle’s view of slavery. As I shall argue in the next section, moral
disagreements among philosophers may be the most promising example of disagreements where it is not so easy to find an asymmetry between the two parties in question.4

2. Philosophical disagreements

For various reasons, it may be thought that if two professional moral philosophers disagree with each other, it will normally be harder to identify an asymmetry between the two sides of the disagreement than in other cases. After all, it may be claimed, professional philosophers have received a rigorous training in philosophy, and were selected for their professional positions through a fiercely competitive process. Given the institutional structure of the philosophical profession, moral philosophers seem to have considerable incentives to think hard and rigorously about these questions. Thus, it seems likely that moral philosophers typically have the cognitive capacities and virtues that will help them to arrive at any truth in ethics that is there to be found; they are likely to be better informed about the relevant non-moral facts than many ordinary thinkers; and in general they are likely to be deploying their cognitive capacities and virtues in relatively favourable circumstances.

If two moral philosophers disagree about a certain moral proposition \( p \), with one philosopher believing \( p \) and the other philosopher believing the negation of \( p \), and there is no way of identifying an asymmetry between these two philosophers, we should conclude that neither of these philosophers knows whether or not \( p \) is true. If these philosophers are also in the best possible position for knowing any truth that could be known on this question, we should conclude in addition that either there is no fact of the matter about whether \( p \) is true, or else it is totally unknowable whether or not \( p \) is true.

---

4 For a version of this argument from moral disagreement among philosophers, see Leiter (2010).
The proponents of this sort of argument must then claim that there are disagreements of this kind about such a wide range of moral questions that we should conclude that either there is no truth about morality at all, or else that all non-trivial moral truths are unknowable.

To defeat this argument, then, it is not necessary to show that there are no moral disagreements of this kind between philosophers. A non-sceptical moral realist could concede that there are indeed some disagreements of this kind about a certain narrow range of moral questions: for example, a non-sceptical realist could concede that it is not always knowable exactly how one should weigh up competing reasons in finely balanced cases.\(^5\) This concession would not imply that there are not many other moral truths that we are capable of knowing.

An authoritative evaluation of this argument, then, would have to be based on an extensive survey of moral philosophers’ opinions. Such a survey would be needed to establish, first, precisely how wide the range of moral questions is that moral philosophers disagree about, and secondly, what proportion of these disagreements can be explained in a way that supports ascribing knowledge to one of the parties to the disagreement. It is clear that a survey of this sort would require extensive empirical investigations.

Rather than undertaking such empirical investigations here, I shall focus on giving a detailed account of the normal methods of moral theory; this account will be given in Section 3 below. Then, in Sections 4 and 5, I will argue that this account makes it seem at least prima

\(^5\) For an argument, inspired by the recent work of Timothy Williamson (2000), for the conclusion that there must be at least some unknowable moral truths, see the following post on the blog PEA Soup <http://peasoup.typepad.com/peasoup/2010/08/there-are-unknowable-moral-truths.html>.\]
facie unlikely that the kind of moral disagreements that exist among philosophers will create a significant problem for non-sceptical moral realism.

In the course of the discussion that follows, I shall touch on two contrasting examples of disagreement between philosophers. The first of these two examples is the disagreement between Kant and most contemporary philosophers over the morality of masturbation and homosexuality. Most contemporary moral theorists would regard such acts as morally permissible, at least so long as they cause no harm, and are done with the mutual consent of everyone involved.\(^6\) In a sharp contrast to this, Kant claims that all such acts are contrary to duty: indeed, along with bestiality, masturbation and homosexuality are categorized as *crimina carnis contra naturam* (crimes of the flesh contrary to nature), which as he puts it, “debase the human condition below that of an animal, and make the human being unworthy of humanity”.\(^7\)

The second example that I shall consider here is the disagreement between contemporary consequentialists and anti-consequentialists about the foundations of ethics.\(^8\) Consequentialists hold that there is some intrinsic feature, which is exemplified to various degrees by the total consequences of acts, such that the ethical status of an act is determined purely by the degree to which the act’s total consequences exemplify this feature. (For

---

\(^6\) For a representative example of such a contemporary theorist, see Corvino (2005).


\(^8\) This is the first moral disagreement mentioned by Brian Leiter (2010).
example, according to utilitarians, this intrinsic feature is the total amount of happiness that exists in the consequences in question.) So, according to consequentialists, if there are two acts whose total consequences exemplify this intrinsic feature to the very same degree, those two acts cannot differ in their ethical status. (For example, according to utilitarians, if the total consequences of two acts involve the very same total amount of happiness, those acts must have the same ethical status.)

Consequentialism is opposed by a rather motley group of moral theorists. So, to fix ideas, let us just focus on one specific group of anti-consequentialists – namely, those theorists who insist that there is crucial ethical difference between doing and allowing. In principle, it seems that two acts could have total consequences that involve exactly the same total amount of happiness, even if one of those acts involved actively doing harm, while the other act involved merely allowing harm. A utilitarian would have to say that these two acts have the same ethical status, while an anti-consequentialist might well say that the act that involves doing harm is morally worse than the act that involves merely allowing harm.

Here then are two striking examples of moral disagreement among philosophers. In the course of the discussion that follows, I shall make some suggestions about the explanation of these disagreements; as I shall try to argue, if these suggestions are correct, they count against the argument that moral disagreement raises a serious problem for non-sceptical moral realism.

3. The methodology of moral theory

What are the methods that lead these philosophers to their moral beliefs? In part, my account of these methods will be thoroughly familiar: I shall endorse the well-known idea that the
methods of moral theory involve pursuing what John Rawls (1971) called “reflective equilibrium”. But I shall also supplement this idea with a more detailed account of the nature of what Rawls called our “considered judgments” – which serve as the primary data that our pursuit of reflective equilibrium sets out from.

Strictly speaking, Rawls only gave an account of how theorists should set about developing a theory of justice. But it seems clear that this account can be generalized so that it covers the whole of moral theory. At all events, Rawls (1971: 46) starts from the assumption that each of us has a “sense of justice”, which involves “a skill in judging things to be just or unjust, and in supporting these judgments by reasons”. Rawls suggests that this skill resembles our ability to judge what counts as a grammatical sentence of our native language – since it involves a capacity for making immediate judgments, in an assured and unprompted way, while there is no limit to the number of judgments that could potentially result from our exercise of this skill.

Among these judgments, Rawls recommends starting out from our “considered judgments”, which are “rendered under conditions favourable to the exercise of the sense of justice, … where the more common excuses and explanations for making a mistake do not obtain” (1971: 47f.). For example, Rawls assumes that considered judgments must be judgments of which one is confident (as opposed to uncertain), and made when one is able to concentrate without distraction on the question at hand (not when one is “upset or frightened”), in situations in which one does not stand to gain or lose depending on how the question is answered.

Provisionally, Rawls suggests, we could think of the goal of this method as formulating a set of general principles that, when conjoined to one’s non-moral factual beliefs, would allow one to derive these considered judgments together with their supporting reasons. However,
even one’s considered judgments may be subject to certain “irregularities and distortions” (1971: 48); so one does not have to find a set of principles from which all one’s considered judgments can be derived in this way. Instead, some of one’s judgments may be revised to conform to the principles, if the principles provide an “intuitively appealing account of [one’s] sense of justice”; one is especially likely to revise one’s judgments in this way if one can also “find an explanation for the deviations that undermines [one’s] confidence in the original judgments.” Once one has formulated intuitively appealing principles that can account for all of one’s considered judgments which have not been revised, one has reached “reflective equilibrium”.

Ideally, moral philosophers should seek what Rawls calls “wide” reflective equilibrium: in constructing this set of principles, one should not only canvas the widest possible array of considered judgments in order to see what sets of principles they can be coherently integrated with; one should also consider every possible set of principles that might conceivably be integrated with one’s considered judgments in reflective equilibrium, together with “all relevant arguments” for these principles.

Since there are many different kinds of “judgments” besides those that issue from a specialized “skill” like one’s “sense of justice”, I shall refer to the mental states that form the initial starting-points for the quest for reflective equilibrium as one’s “intuitive moral judgments”. A central question for moral epistemology is where these “intuitive moral judgments” come from. I have offered an answer to this central question in some of my earlier work (Wedgwood 2007a: ch. 10). To fix ideas, I shall assume here that my answer to this question is correct. For the purposes of this essay, however, it does not matter whether this answer is correct in every detail. Since our goal here is just to demonstrate that non-sceptical moral realists are not completely at a loss to explain disagreements among moral
philosophers, all that matters is that my account of where our intuitive moral judgments come from is both reasonably plausible in itself and compatible with moral realism.

I propose that an intuitive judgment of this kind is a moral belief that we hold, not because we have inferred it from any other beliefs, but simply because, when we consider the proposition that is the content of the belief, we have an immediate inclination of a certain kind to believe that proposition. This immediate inclination to believe the proposition could itself be called the “moral intuition” underlying this intuitive moral judgment. But this raises the further question: Where do these moral intuitions come from?

In a nutshell, I propose that our moral intuitions come from some of our emotional dispositions – specifically, from our dispositions towards moral emotions. There are two main ways in which moral intuitions can arise from these emotional dispositions. First, in some cases we are confronted with a situation that we believe to be actual, and respond to our beliefs about this situation by having an emotional reaction – and this reaction inclines us to believe that the situation has the moral features that would make this emotion appropriate. For example, you might read the newspaper and as a result form certain ordinary non-moral beliefs about how a certain politician has behaved, and respond to these beliefs by feeling outraged at the politician’s behaviour – in which case your response would incline you to believe that the politician’s behaviour really was morally outrageous, and this inclination would be a what I am calling a “moral intuition”.

As Sarah McGrath (2004) has plausibly argued, there are also apparent cases in which we gain moral knowledge by perception; I would interpret these cases as involving this first kind of moral intuition.
Secondly, in some other cases, we merely imagine a situation of some general kind, without positively believing that there is any actual situation of this kind; in these cases, we respond to our imagining of this situation with a simulation of the reaction that our emotional dispositions would produce in response to really believing ourselves to be confronted with a situation of this kind – and this response would incline you to believe that, at least normally and ceteris paribus, situations of the imagined kind have the feature that would make that emotion appropriate. For example, you might imagine a “trolley problem”, of the sort that has been discussed by Philippa Foot (1978) and Judith Thomson (2008) among others, and your response to this imagining might involve a simulation of a certain emotional reaction, which would incline you to form the corresponding moral belief about situations of this kind; and this inclination would also be a moral intuition.¹⁰

Some philosophers may be sceptical of the suggestion that our moral intuitions arise from our emotional dispositions in these ways. Some of them may try to argue that these intuitions are more like our purely intellectual intuitions of the self-evident truths of mathematics or the like. But the evidence of recent empirical work in moral psychology tells strongly against such an intellectualist view of moral intuitions, and in favour of the more sentimentalist view that I have articulated here.¹¹

¹⁰ This account of our moral intuitions can also explain why our capacity for such intuitions is productive, just as Rawls hinted with his analogy between our moral capacity and our ability to judge which sentences of our native language are grammatical. Our emotional dispositions have an inherent generality to them: they yield reactions to situations that the thinker has never considered before.

¹¹ As psychologists such as Jonathan Haidt (2001) have stressed, there is much empirical evidence for the conclusion that emotions play a very extensive role in moral thinking.
Other philosophers may suspect that the view that our moral intuitions arise from our emotional dispositions is in tension with a realist conception of moral truth. In my opinion, this is not so.\textsuperscript{12} Moral realists can consistently hold that there is an objective fact of the matter about whether or not a given emotional reaction is appropriate to the situation in question; and if the emotional reaction is appropriate, then the situation must have a corresponding moral feature that makes that reaction appropriate, and the moral proposition that the emotional reaction inclines the thinker to believe will be true.

Moreover, moral realists can also hold that it is rational for thinkers to presume, at least until they acquire any special evidence to the contrary, that their emotional reactions are appropriate; and so it is also rational for them to treat the intuitions that arise from their emotional dispositions as defeasible reasons in favour of the corresponding moral beliefs. For the purposes of the present discussion, I shall simply assume that this sentimentalist conception of moral intuitions is both correct and compatible with moral realism.

This account of the methods of moral theory makes it clear that there are three possible sources of error in moral theory:

(i) The moral intuitions that serve as the starting-points for moral theorizing may be misleading;

(ii) The process of searching for reflective equilibrium may lead these theorists astray;

(iii) The non-moral beliefs that are assumed by these moral theorists may also be false.

For our purposes, we can ignore this third source of error here: as we saw with the example of Aristotle’s views of slavery, if the fact that a philosopher holds a false moral belief is adequately explained by the philosopher’s holding a false non-moral belief, then no special...

\textsuperscript{12} For a defence of this opinion, see Wedgwood (2007a: ch. 10).
problems for non-sceptical moral realism will ensue. So, the sources of error that we need to consider are the first two: (i) the philosopher’s initial moral intuitions; and (ii) the process of searching for reflective equilibrium.

4. Sources of error in moral theory (i): Initial moral intuitions

If our moral intuitions do indeed arise from our emotional dispositions in the way that I have proposed, we should expect our moral intuitions to reveal a significant degree of interpersonal variation. Moreover, our moral intuitions are just as susceptible of being moulded by cultural influences as our emotional dispositions. Where these individual or cultural differences lead thinkers to form incompatible moral beliefs, the intuitions on at least one side of this difference must be misleading intuitions – that is, intuitions arising from emotional dispositions that are manifested in inappropriate emotional reactions. Such misleading intuitions are in some ways similar to perceptual illusions or hallucinations.

If our moral intuitions arise from our emotional dispositions in the way that I have described, there are, in general, as many ways of explaining why someone has misleading moral intuitions of this kind as there are of explaining why someone has a disposition towards inappropriate moral emotions. Psychologists have studied our emotional dispositions in great detail; when supplemented with a moral theory about when these emotions are appropriate, these psychological studies can enable us to explain why a person has a disposition towards an inappropriate emotional reaction (as well as why they have dispositions towards more appropriate emotional reactions).

Philosophers are no more immune to misleading intuitions than anyone else; at best, they have a slighter better chance of discovering that some of their intuitions are misleading by the
method of pursuing reflective equilibrium. But this method itself relies, in a holistic way, on one’s total set of intuitions; and so it is not surprising that many philosophers will end up with mistaken moral beliefs, given that some of the intuitions that they start out from may be misleading in this way.

Consider a case in which two philosophers disagree, and the best explanation of the disagreement appeals to the different intuitions that these philosophers start from. It may be that in this case, one philosopher’s intuition arises from a disposition towards inappropriate emotional reactions, while the other philosopher’s intuition arises from a disposition towards more appropriate emotions. This seems to be the kind of asymmetry between the two philosophers that could explain why one of the philosophers has knowledge (in spite of the superficial similarity between this philosopher and the other philosopher who believes a false proposition about the matter in question). So disagreements of this kind do not raise any problems for non-sceptical moral realism.

It seems plausible to me that the disagreement between Kant and the contemporary moral theorists about the ethics of homosexuality and masturbation is at least in part a disagreement of this kind. Kant’s discussion of acts of these types uses surprisingly strong language: he describes these acts, not just as “base” (niedrig), but as the “most contemptible acts” (das verächtlichste) that a human being can commit; as he says, these acts are “unmentionable” because even mentioning these acts “causes disgust” (wenn man sie nennt ein Ekel verursacht wird). This language seems clearly to reveal that Kant has a distinctive emotional response when thinking of these acts: he responds with a feeling of revulsion and disgust. Given my account of moral intuitions, this supports the interpretation that Kant has the intuition that masturbation and homosexual acts are simply foul and disgusting.
Psychologists have studied these emotional reactions in which certain sexual acts can seem “impure” to some people – even if those who participate in those acts consent to them, derive only pleasure from them, and are not harmed by them in any tangible way.13 People who have these emotional reactions are often inclined to believe that the acts in question are not just disgusting for them to contemplate, but somehow intrinsically disgusting in themselves – which often also inclines these people to judge these acts to be morally wrong. Now it seems to me plausible that the correct ethical theory will imply that these judgments are false, and so these emotional reactions are to that extent misleading. So, by putting these psychological studies together with the most plausible ethical theory, we can explain why Kant has misleading moral intuitions on this topic.

However, there is a further complication with our intuitions that is especially relevant to explaining philosophical disagreement. As I have explained, when a moral intuition arises from an emotional disposition of the right kind, then the moral proposition that is the content of the intuition will be true. However, even in these cases, when the content of the intuition is true, it is usually not an explanatorily fundamental truth. The moral truths that are reflected in our emotional dispositions include many truths that depend in part on the empirical facts about the actual world that we find ourselves in. We react in horror at the mere thought of killing people; but arguably this is only because of the empirical fact that in our world, when people are killed they are almost always seriously harmed as a result. If we lived in a world in which it was impossible to kill anyone except those who wish to be killed in order to be released from an excruciating terminal illness, then the idea of killing people would not arouse similar emotional reactions.

13 See for example Cannon et al. (2011).
Some philosophers – most notably, Christopher Peacocke (2004) – have argued that all moral thinkers are guided in their moral thinking by an implicit grasp of the fundamental *a priori* necessary principles of morality. According to this view, ordinary thinkers make moral judgments by implicitly drawing *inferences* from these fundamental necessary principles, together with ordinary empirical information about the relevant circumstances. As I have argued elsewhere, this view of moral thinking has implications that are not credible (see 2007b). So it seems that normally, even when the content of one of our moral intuitions is true, it is not an explanatorily fundamental necessary truth, but a *contingent* moral truth instead.

It may be that each of these contingent moral truths can, at least in principle, be ultimately explained on the basis of a conjunction of non-moral empirical truths and some ultimate necessary principles. However, these ultimate necessary principles are not immediately revealed in our ordinary intuitions. In that sense, when these intuitions are true, what makes them true is normally a mixture of the purely moral and the purely non-moral; and the nature of this mixture is not transparent to the ordinary thinker who has the intuition. In short, in these normal cases, the content of our intuitions are mid-level contingent moral truths, rather than the explanatorily fundamental necessary truths.

With respect to mid-level moral truths of this kind, there actually seems to be a fairly impressive amount of *agreement* among contemporary philosophers. The truths about prudential reasons are hardly controversial at all. Almost everyone agrees that people normally have prudential reason to avoid pain, to look after their health and financial security, and to stay alive. Certain central moral truths are equally widely agreed. Almost everyone agrees that we should normally keep our promises, refrain from killing and stealing, be grateful to those who have been kind to us, and so on. Typically, the areas where philosophical disputes arise concern either (a) cases that are relatively peripheral to most
people’s moral sensibility (such as the ethical status of human foetuses and non-human animals), or (b) the ultimate principles that explain all moral truths.

For this reason, it seems that the propositions that are the contents of most of our moral intuitions can be explained equally well by a wide range of ultimate ethical theories; and this seems particularly true of our intuitions about all the normal cases that we most commonly encounter in everyday life. So, to distinguish between these ethical theories, it is often necessary to canvas our intuitions about some rather abnormal cases.

For example, consider the kind of case that can unequivocally separate all forms of consequentialism from all theories that ascribe intrinsic ethical significance to the distinction between doing and allowing. It would have to be a case like the following. Consider two agents – you and me – who are qualitatively identical in all ethically relevant respects.

Suppose that there are three acts available to you: either (a) you kill an innocent person, Victim-1, or (b) you allow me to kill a qualitatively identical person, Victim-2, or (c) you prevent me from killing Victim-2 at a terrible cost to yourself. Similarly, suppose that three precisely similar acts are available to me: either (a’) I kill Victim-2, or (b’) I allow you to kill Victim-1, or (c’) I prevent you from killing Victim-1 at a terrible cost to myself. Moreover, in this situation, if you were to kill Victim-1, I would allow you to do so (and I would not kill Victim-2); and if you were to allow me to kill Victim-2, I would kill Victim-2 (and you would not kill Victim-1). So the only difference between the total consequence of (a) your killing Victim-2 and the total consequence of (b) your allowing me to kill Victim-1 is that you and I, and Victim-1 and Victim-2, have switched places; otherwise the two consequences are exactly the same. So every form of consequentialism would have to accept that in this case
the agent-neutral value of the two consequences is also exactly the same. According to consequentialism, then, the two acts (a) and (b) cannot differ in their ethical status.\textsuperscript{14}

On the other hand, proponents of the doing / allowing distinction would most likely regard these two acts as differing significantly in their ethical status: according to these theorists, your allowing me to kill Victim-2 could be \textit{permissible} (given that the only way in which you could save Victim-2’s life is by preventing me from killing Victim-2 at a terrible cost to yourself), while your killing Victim-1 yourself would be \textit{wrong} or \textit{impermissible}. (For example, it seems clear that Judith Thomson (2008) would draw this conclusion about cases of this sort.)

However, it is clear that this is a fairly unusual case – a case where the total consequences of your doing harm and of your allowing harm are exactly the same except that the positions of some qualitatively identical individuals (you and me, and Victim-1 and Victim-2) are switched around. As it seems to me, there is a question about whether one can know a moral proposition by relying solely on an intuition about an unusual case of this sort.

Our intuitions arise from our emotional dispositions, which have evolved (through biological evolution and through individual learning and development) to deal with the everyday situations that we normally find ourselves in. For this reason, our intuitions about such unusual situations are in a sense \textit{less clear} than our intuitions about more ordinary situations: it is harder for us to think about these unusual situations in a way that is reliably sensitive to how our emotional dispositions really incline us to react to these situations. There seem to be two reasons for this.

\textsuperscript{14} For an example of this consequentialist view of such cases, see for example Norcross (2008) and Unger (1996).
First, since these unusual situations are so different from the ones that we normally
encounter, it is harder for us to respond to our imagining of such an unusual situation with an
accurate *simulation* of the emotional reaction that we are disposed to have to actually
believing ourselves to be confronted with an actual situation of this kind.

Secondly, every mental disposition involves being such that there is a certain range of *normal
cases* in which one manifests the disposition, by reacting to the relevant stimulus with the
relevant response. But it seems plausible that the kind of “normality” that is in question here
comes in *degrees*. Strictly speaking, then, the connection between the disposition and its
manifestations should be characterized in *probabilistic* terms: the less normal a case is, the
lower the chance that the disposition will be manifested in the case. A case in which one
believes oneself to be confronted with a highly unusual situation, of the sort that one is
imagining in having one of the intuitions that we are currently considering, seems to be a
relatively abnormal case. So the chance that one will respond appropriately to believing
oneself to be confronted with an unusual situation of this sort may be less than the chance of
responding appropriately to more ordinary cases.

So, intuitions about such unusual cases seem likely to be less reliable than intuitions about the
more ordinary cases. Even if they are only slightly less reliable, they are at least closer to the
line that separates cases of knowledge from cases of mere luckily true belief. For this reason,
an intuition about an unusual case of this sort may not always be reliable enough to serve as
the sole basis for knowledge of the corresponding moral proposition.

There is a way of trying to have a more reliable belief about whether or not an ultimate
ethical theory like consequentialism is true. One can supplement bare reliance on one or two
intuitions with a more holistic assessment of these rival theories – by searching for reflective
equilibrium. This leads us to the second potential source of error in moral theory.
5. Sources of error in moral theory (ii): The search for reflective equilibrium

According to Rawls, the task of ethical theory is to develop a set of principles from which one can “derive” the contents of those moral intuitions that are accepted in “reflective equilibrium”. I propose, however, that the set of principles that ethical theorists should seek are those that provide the ultimate explanation of all other moral truths. In this sense, the method of ethical theory is that of an inference to the best explanation. The ethical theorist must canvas as large a number as possible of moral intuitions, and then seek the best explanation for the moral propositions that are the contents of those intuitions that survive the quest for reflective equilibrium.

Judging which of the many possible ethical theories is the best explanation of these moral propositions requires a complicated holistic judgment, comparing the pros and cons of all the different possible theories. The task is complicated by the fact that as I have just emphasized, our intuitions reflect a non-transparent mixture of empirical non-moral information and fundamental moral principles. For all these reasons, then, the project of canvassing a sufficiently large body of moral intuitions, and finding the best explanation of the contents of those intuitions that survive the quest for reflective equilibrium, is a challenging and potentially confusing task.

In general, it is clear that knowing the ultimate principles of ethics is just as hard as knowing the fundamental explanations of other large-scale phenomena where we have a large and complicated body of data that does not point unequivocally in the direction of any one theoretical explanation. Consider, for example, history. Admittedly, some of the questions of history are simple (‘Did the South lose the American Civil War?’, ‘Was Julius Caesar assassinated?’). But questions about the ultimate explanation of large-scale phenomena are much harder. For example, there has been a vigorous debate among British historians, for at
least the last fifty years, about whether the Protestant Reformation in 16th-century England was fundamentally a bottom-up process, in which Protestant ideas sparked off a mass religious movement that the authorities were forced to try to manage as best they could, or whether it was principally a top-down process, imposed by the country’s political elite on a largely unwilling or indifferent populace (see Duffy 1992). There is no sign that this debate is about to be resolved within the foreseeable future.

Equally, we might compare debates elsewhere in philosophy – for example, in metaphysics and philosophy of language and the like. These debates also show little sign of being about to be resolved in the immediate future, but it would be unwarranted to conclude either that there is no truth of the matter about philosophy, or that absolutely all philosophical truths are wholly unknowable.

Thus, the conclusion supported by these considerations is not a sweeping moral scepticism of the sort that we considered at the outset of this paper, but a much more urbane and limited form of scepticism: in effect, the thesis that the most fundamental principles of ethics are significantly harder to know than less fundamental moral truths.

For this reason, it seems to me that if moral philosophers were perfectly rational, they would not have complete confidence any particular ethical theory at all. They would have a mere partial degree of belief instead. So perhaps moral philosophers, if they are rational, would not have complete confidence either in consequentialism, or in the claim that the distinction between doing and allowing has the kind of ethical significance that anti-consequentialists have claimed; they would have only a partial degree of belief instead. This partial degree of belief might be too low for the belief to count as “knowledge”, but this point will only apply to very general ultimate principles of this sort. Moral philosophers might still know a great many less ultimate moral truths.
Indeed, it may be that moral philosophers do not in fact have quite as much confidence in the ethical theories that they defend as they sometimes appear to have. Philosophy has, ever since the days of Socrates, had a culture of adversarial debate. So it is common for philosophers to adopt a position that they defend vigorously in debate (while also vigorously criticizing the rival philosophical positions that are incompatible with their own). But it does not follow that these philosophers are \textit{totally convinced} of the truth of their own position (or even of the falsity of the rival positions). Philosophers may only have a weak level of belief in the position that they are defending. That is, they may only regard their favoured theory as being \textit{slightly more likely} to be true than any of the rival theories that deserve to be taken seriously.

Let me illustrate this point with an artificially precise example. Suppose that there are three theories that I believe to deserve to be taken seriously about a certain philosophical issue – call these three theories T1, T2, and T3. Now suppose that I assign a 30% probability to T1 and to T2, and a 40% probability to T3. Then I regard T3 as more likely than any of its rivals, even though I believe that the probability of its being true is only 40%, while the probability of its being false is a full 60%. Still, even this low level of belief may make it rational for me to pursue a strategy of developing this theory T3 and defending it in debate, since this strategy may be a good way to advance philosophical understanding.

There may even be something of a convention in philosophy to express opinions with an air of greater confidence than one really feels. This may serve the goal of advancing understanding by promoting vigorous debate. Such debate can advance philosophical understanding even if it does not give a conclusive answer to all the ultimate theoretical questions. Such debate can provide illuminating answers to a host of \textit{smaller} questions: What are the most promising forms that each kind of answer to these ultimate questions can take? What are the strong and the weak points in each of these answers? And so on. At all events, we cannot infer from the fact that philosophers have a culture of vigorous debate that
philosophers “disagree” with each other in the strong sense of being fully convinced of incompatible positions.

One might complain that this only explains why we can never be certain of the correct answer to these ultimate philosophical questions. It does not explain why philosophers hold false beliefs about these matters. As I have already indicated, I do not think that philosophers – at least if they are rational – are strongly convinced of the truth of their favoured answer to these ultimate questions. So the problem is just to explain how it could be that I assign a 40% probability to the philosophical theory T3, and only a 30% probability to the rival theory T1, if in fact the true theory is T1 and not T3. But if these issues are just intrinsically difficult and obscure, for the reasons that I described above, it is not surprising that different philosophers give slightly different probability assignments to the various different theories in this domain.

Even if it were clearly established that many philosophers really are firmly convinced of their favoured answer to these ultimate questions, this conviction can be explained by a familiar kind of irrationality. Philosophers can easily “fall in love” with the theory that they are defending, and so come to believe the theory with greater confidence than they are really entitled to.

Indeed, the institutions of academic philosophy make it easy to explain why this sort of irrationality might arise. On the easier questions, philosophers are penalized for making mistakes, by getting lower grades while they are students, and by failing to achieve success after they join the profession. But given the extreme difficulty of knowing the right answer to the harder questions (including these ultimate questions of ethics), there is no institutional penalty for giving what are in fact incorrect answers to these harder questions. Instead, philosophers are rewarded for developing original answers to these harder questions. It
probably helps to motivate a philosopher to put in the hard graft that is required to develop an original answer to one of these harder questions if the philosopher has a fairly high degree of confidence in the correctness of this answer. So it is not hard to explain why some philosophers have a somewhat higher degree of confidence in the correctness of their favoured theory than they are rationally entitled to.

As I have noted, these phenomena may be able to explain the debates between consequentialists and the proponents of the doing / allowing distinction. This explanation is certainly quite compatible with a realist conception of moral truth. It does indeed imply that on these foundational questions, we should be cautious and should not have full confidence in any particular account of the fundamental principles of morality. In this way, this explanation of this disagreement is at best modestly non-sceptical. Nonetheless, this explanation is clearly compatible with the view that we can know a great many of the less fundamental truths of morality. So this explanation is compatible with non-sceptical moral realism.

Here too, then, we have no difficulty in explaining the sort of moral disagreement that exists among philosophers in a way that is compatible with non-sceptical moral realism. The familiar argument that non-sceptical moral realism conflicts with the most plausible explanations of moral disagreement seems on reflection to be much less cogent than it is often claimed to be.¹⁵

¹⁵ Some of the ideas for this paper were presented on the National Humanities Center’s blog site, On the Human, in response to a post by Brian Leiter (2010). I am grateful to Leiter and the other commentators on that site for helpful comments. Prior to the Challenges conference at Purdue, I received extremely valuable written comments from Patrick Kain; and at the
References


conference itself, I received many further helpful comments, both from my commentator, John Pittard, and from the other participants.


