

The Moral Evil Demons

Ralph Wedgwood

Moral disagreement has long been thought to create serious problems for certain views in metaethics. More specifically, moral disagreement has been thought to pose problems for any metaethical view that rejects *relativism*—that is, for any view that implies that whenever two thinkers disagree about a moral question, at least one of those thinkers' beliefs about the question is not correct. In this essay, I shall outline a solution to one of these problems. As I shall argue, it turns out in the end that this problem is not really a special problem about moral disagreement at all: it is a general problem about disagreement as such. For this reason, in the later sections of this essay, I shall turn to some general questions in epistemology, about the epistemic significance of disagreement.

1. The problem of the moral evil demons

There are several different ways in which relativists have argued that moral disagreement poses a problem for their opponents. For example, relativists such as Gilbert Harman (in Harman and Thomson 1995) have argued that relativism gives a *better explanation* of the sort of moral disagreement that exists than any rival view. According to these relativists, the best explanation of this sort of disagreement involves the hypothesis that both sides of the disagreement are in their way correct—whereas no equally good explanation involves the hypothesis that at least one side of the disagreement holds a belief that is not correct. Whether or not these relativists are right to argue this is a complicated empirical question. As several opponents of relativism, such as Judith Thomson (in Harman and Thomson 1995), have argued, it seems that there are in fact a great many social and psychological mechanisms that could perfectly well explain why we would end up with seriously distorted views about many moral questions.

At all events, it is not this problem that I shall focus on here. I shall suppose that the anti-relativist can give an explanation of the existence of moral disagreement that is at least as good as the explanation that is offered by the relativist. I shall also suppose that the anti-relativist has succeeded in developing a plausible moral epistemology, according to which it is rational for one to form moral beliefs on the basis of one's moral intuitions, at least so long as those moral intuitions have a reasonable degree of coherence with one's overall set of moral beliefs.¹ Even if we grant all these assumptions to the anti-relativist, a further problem seems to arise: once we *learn* about the sort of disagreement that actually exists, why doesn't this information remove any justification that we might previously have had for the relevant moral beliefs? Why doesn't the information about all the moral disagreement that exists force us into a thoroughgoing *scepticism* about our moral beliefs?

This problem arises most clearly in cases of moral disagreement between two thinkers who are *equally rational*, and *equally well informed about the non-moral facts*. Obviously, there are many cases where moral disagreement is explained by the

¹ For an example of a moral epistemology of this kind, see Wedgwood (2007, Chap. 10)—although for the purposes of this discussion, we do not need to presuppose the exact details of that account of what moral intuitions are, or where they come from.

fact that one party to the disagreement (or perhaps even both parties to the disagreement) are less well informed about the non-moral facts than they might have been: one party to the disagreement might simply be ignorant of certain non-moral facts; or one party might actually have mistaken or erroneous beliefs about these non-moral facts. Similarly, there are also many cases where moral disagreement is explained by some sort of procedural irrationality on one side or the other. For familiar reasons, bias and self-interest are particularly likely to cause self-deception about moral questions; as a result, people often persist in their moral beliefs, in spite of the fact that their own stock of beliefs and other mental states would have motivated them to abandon those moral beliefs if they had reflected more rationally about the question.

However, there is no obviously compelling reason why we should deny the possibility of moral disagreements that are not of these kinds.² In the absence of any such obviously compelling reason, we should assume that it is possible for there to be moral disagreements that are not explained either by irrationality or by any lack of non-moral information on either side.

In what follows, I shall suppose that this assumption is correct: it is possible for there to be moral disagreements of this kind—that is, disagreements in which both parties to the disagreement are forming and revising their beliefs in procedurally quite rational ways, and neither side holds their belief because of any error or ignorance about the purely non-moral facts. Indeed, it may even be that some disagreements of this sort are *actual*. For example, some thinkers believe that it is morally wrong for people to eat meat (unless those people have to eat meat in order to stay alive and well), while others believe that eating meat simply for the pleasure of doing so is perfectly permissible. This disagreement may not be due to any irrationality, or to any non-moral error or ignorance, on either side.³ There may be many other such disagreements: for example, there are all the disagreements about sexual morality (for instance, about whether or not there is something morally inferior about homosexuality compared to heterosexuality); there are political disagreements about what forms of liberty or equality are important, and why; there is the disagreement about the moral status of early human foetuses and embryos, and the disagreement about whether the intrinsic value of species diversity and thriving natural ecosystems gives us any moral duties to respect this value; and there are many disagreements about how to balance different moral values or reasons—for instance, how to balance individual rights against collective security, individual autonomy against social order and cohesion, and so on.

If moral disagreements of this kind are not explained either by procedurally irrational reasoning or by non-moral error or ignorance, what does explain these disagreements? In such moral disagreements, it seems that the two parties hold their incompatible beliefs, not because of procedurally irrational reasoning or non-moral error or ignorance, but simply because the two sides have sufficiently different *pre-theoretical moral intuitions*, which lead them to believe different fundamental moral

² Richard Boyd (1988, 221) advocates a view that he calls the “rational supervenience” of moral beliefs on non-moral beliefs, which denies the possibility of any moral disagreements that do not involve either irrationality or disagreement about the non-moral facts—except in a “few cases” where either bivalence fails (so that in fact neither side in the disagreement is determinately correct or incorrect), or else “explanations in terms of nonculpable inadequacies in methodology or theoretical understanding are readily available” (222). However, it is not clear what reason Boyd has for accepting these claims.

³ For a longer and more careful argument for this point, see Richard W. Miller (1992, chap. 1).

principles. (For example, vegetarians may have the moral intuition that any creature with the capacity for pain and suffering has the kind of status that makes it impermissible to kill it just for the pleasure of eating it—while carnivores may think that we have much less powerful reasons to refrain from killing non-rational animals than we have to refrain from killing animals, like human beings, who have either the capacity for rational thought or at least the potential to develop this capacity.)

As I noted above, my goal here is to solve a problem that moral disagreement creates for those metaethical views that oppose relativism—that is, for those views that imply that whenever two thinkers disagree about a moral question, it is impossible for both thinkers to be right. So I shall assume here (at least for the sake of argument) that some metaethical view of this general kind is correct. Given classical logic, it follows that whenever two thinkers disagree, at least one of the parties has a false or mistaken belief about the question in dispute. So if the disagreement is due to the two parties' having sufficiently different pre-theoretical intuitions, at least one of the parties must have had *misleading* pre-theoretical intuitions, which has led them to a false and mistaken belief about the question. (I assume here that there need be nothing *irrational* about having such misleading moral intuitions—just as there need be nothing irrational about undergoing a hallucination or optical illusion.)

Now in some cases of this sort, the pre-theoretical moral intuitions of one of the two parties may contain some sort of incoherence that is not present in the intuitions of the other party. In that case, even though each of the two parties will base their thinking about this issue on their pre-theoretical moral intuitions, it may be fairly easy for the party whose belief is in fact mistaken to discover their mistake by means of this kind of thinking. In some other cases, however, the misleading pre-theoretical moral intuitions of this party to the disagreement may be relatively *systematic*: that is, although these intuitions are in fact misleading, they also form an overall set that is no less coherent than the intuitions of the other party.

In a case of such systematically misleading pre-theoretical intuitions, even though one of the two parties has an incorrect or mistaken belief, it seems that ordinary moral reasoning will be incapable of leading the believer to discover this mistake. It seems inevitable that any further reflection on the part of this believer will be based on the same systematically misleading pre-theoretical intuitions. Since these intuitions contain no incoherence that would alert the believer to his mistake, it is hard to see how further reflection based on these intuitions could lead the believer to correct his mistake. Something has caused the thinker to have the systematically misleading initial intuitions that he has—his upbringing, or his culture, or his character, or something like that. Whatever it was, I shall call it a “moral evil demon”—something that causes moral error in a way that makes that error undetectable by ordinary means to the one who is deceived.

There is a striking difference between these moral evil demons and their more famous cousins, the Cartesian evil demons, who deceive their victims by giving them systematically misleading sensory experiences. The Cartesian evil demons are creatures of philosophical fantasy; they are not to be found in the actual world. Of course, hallucinations and optical illusions do occur. But when they do occur, there is usually some sort of incoherence in the content of one's experiences so that it is possible to avoid being led into any mistaken beliefs. In real life, sensory hallucinations are never so systematic that they cannot be detected by the ordinary methods of empirical thinking. On the other hand, there is a good chance that the moral evil demons are *actual*. No doubt many moral disagreements are explained by procedurally irrational thinking on one side or the other, or by error or ignorance

about relevant non-moral facts. But it seems that there are some disagreements that are more plausibly explained by people's pre-theoretical moral intuitions; and in some of these cases, it is not clear that the intuitions of either of these disagreeing parties contains any sort of internal incoherence that would lead them to change their view on further reflection. In these cases, then, people's pre-theoretical moral intuitions are leading them astray, in a way that resists correction by ordinary moral thinking: that is, a moral evil demon has been at work.

This suggests a different argument for scepticism from the argument that is based on the mere *possibility* of an evil demon. It seems that we all have strong reason to suspect that we live in a world in which moral evil demons are *actually* at work. So what entitles you to any confidence that *your* moral intuitions have not been led astray by such a moral evil demon? What reason do you have to think that you are immune to their malign influence? But if you think that there is a significant chance that your own moral beliefs have been distorted by the influence of such a moral evil demon, surely you should entertain some very serious sceptical doubts about your moral beliefs?

2. Sidgwick's principle

So far, this is still a very rough and impressionistic statement of this argument for scepticism about our moral beliefs. We need to lay out this argument in more explicit detail. It might seem that this argument for scepticism about moral belief has the same structure as the following: Suppose that you are in a prison where you have strong reason to suspect that prisoners are *actually* routinely anaesthetized in their sleep and then have their brains removed and placed in vats. Surely this should lead you to entertain very serious doubts about your ordinary perceptual beliefs. It may seem that in just the same way, once we become aware that we have strong reasons to suspect that moral evil demons are actually at work, we should entertain very serious doubts about our moral beliefs—indeed, perhaps we should even suspend judgment completely about the large parts of our moral thought that seem likely to be subject to disagreements of this sort.

In fact, however, it seems doubtful whether the argument from the probable actual existence of moral evil demons to a sceptical conclusion about our moral beliefs can be quite the same as the seemingly analogous argument in the case of those who are held in a prison where prisoners routinely have their brains placed in vats. In the latter case, you would have a compelling reason to think that there was a nearby possible world in which you used exactly the methods that you actually use to form the very same perceptual beliefs that you actually form (such as 'There is a prison guard dressed in blue standing in front of me'), in which those perceptual beliefs are false. That is, in the prison case, you have compelling reason to regard your perceptual beliefs as unsafe, and as formed by means of a method that is unreliable in the circumstances.⁴

On the other hand, it is not clear that even if you have compelling reason to think that you live in a world in which moral evil demons are at work, it necessarily follows that you have compelling reason to think that your moral beliefs are unsafe, or formed by means of a method that is unreliable in the circumstances. Even if there are

⁴ For this notion of safety, see Williamson (2000, 123-8); for the idea of a method that is unreliable in the circumstances at hand, see Wedgwood (2002a).

moral evil demons at work in the actual world, it does not follow that there is any nearby possible world in which *your* moral beliefs are false. Suppose that you believe the proposition ‘It is permissible for human beings to eat humanely killed chickens, purely for pleasure’. If this proposition is true, then it is presumably true at *all* worlds, except perhaps for some worlds that are very remote from the actual world indeed (such as worlds in which chickens are as intelligent as four-year-old human children, perhaps). So if it is true, there is no nearby world in which it is false—and *a fortiori* no nearby world in which it is false and you believe it.

According to a stronger conception of what it is for a belief to be “safe” (or formed through a “reliable” method), for one of your beliefs to be safe it is not enough that there is no nearby world in which you believe that very proposition and it is false, but also no nearby world in which you believe any sufficiently similar proposition as a result of a sufficiently similar method, and the proposition believed is false. But it is still not clear that even if you do have a compelling reason to think that you live in a world in which moral evil demons are actually at work, it follows that any of *your* moral beliefs are unsafe. Perhaps your upbringing, your brain chemistry, and the cultural influences to which you have been subject have all been thoroughly salutary and benign. A world in which you were instead affected by a moral evil demon instead of these benign and salutary influences would be a world in which your whole life was significantly different from how it actually is; and such a world would presumably not be one of the relevant nearby worlds. So even on this stronger definition of safety, we still do not have an argument for the conclusion that you have a compelling reason to doubt the safety of your moral beliefs.

However, it still seems plausible that a sceptical argument of some kind could be developed out of the reasons that I have canvassed for thinking that we live in a world in which moral evil demons are actually at work. I propose that the reason why this seems so plausible is as follows. Once one recognizes that moral evil demons may be actually at work, this recognition awakens the suspicion that one’s own moral intuitions may have distorted by such a moral evil demon; and this leads us to think that this suspicion must be dismissed on some basis that is wholly *independent* of the moral intuitions in question. However, the very nature of the moral evil demons ensures that there can be no such fully independent basis for dismissing the suspicion that one’s moral intuitions may have been distorted by a moral evil demon: it seems that any argument for the reliability of one’s moral intuitions would itself have to depend on one’s moral intuitions, and so would fail to count as an independent basis for dismissing this suspicion.

Why should the actual existence of moral evil demons give rise to any such suspicion? Of course, one way in which it might do so is because it makes salient the *possibility* that one is somehow deceived; but this way of arousing doubts gives no special role to the evidence that one has that these moral evil demons are not just possible but *actual*. So I propose that the way in which the actual existence of moral evil demons gives rise to sceptical doubts essentially involves a principle about *actual disagreement*—roughly, the principle that whenever one believes a proposition *p*, and learns that some other thinker disbelieves *p*, then one should suspend judgment about *p* unless one has some independent grounds for regarding the other thinker as less likely to be right about *p* than one is oneself.

Versions of this principle have been defended by a number of philosophers. One prominent early example is Henry Sidgwick (1907, 342):

if I find any of my judgments, intuitive or inferential, in direct conflict with a judgment of some other mind, there must be error somewhere: and if I have no more reason to suspect error in the other mind than in my own, reflective comparison between the two judgments necessarily reduces me to state of neutrality.

There are also similar claims in other works of Sidgwick (2000, 168):

I suppose that the conflict in most cases [of philosophical controversy] concerns intuitions—what is self-evident to one mind is not so to another. It is obvious that in any such conflict there must be error on one side or the other, or on both. The natural man will often decide unhesitatingly that the error is on the other side. But it is manifest that a philosophic mind cannot do this, unless it can prove independently that the conflicting intuitor has an inferior faculty of envisaging truth in general or this kind of truth; one who cannot do this must reasonably submit to a loss of confidence in any intuition of his own that thus is found to conflict with another's.

Many questions could be raised about how to interpret these passages. But I shall ignore these questions here. Instead, I shall just assume that the underlying principle behind all these claims is the following:

If you have a belief about a (first-order) question, and then acquire the (higher-order) information that another thinker disagrees with you about that question, you are rationally required to suspend judgment about that (first-order) question, unless you have *independent* grounds for thinking that the other thinker is less reliable about that question than you are yourself.

Just to have a label, I shall refer to this principle as “Sidgwick’s principle”.

Many more recent philosophers have made claims that seem very similar to Sidgwick’s principle. Thus, Adam Elga (2007) claims that whenever you learn that another thinker attaches a different credence to a proposition p from the credence that you attach to p , you should adjust your credence to what Elga calls your *prior conditional credence* in p , conditional on the assumption that the other thinker disagreed with you in the way in which he actually does. As Elga explains, by referring about your “prior” conditional credence in this way, he means a conditional credence that is *prior to and independent of* any reasoning that led to your precise view about this particular proposition p . Broadly similar views have also been advocated by David Christensen (2007) and Richard Feldman (2006).

If Sidgwick’s principle is correct, then given that there is disagreement about a large number of moral propositions, you should suspend judgment about those moral propositions unless you have independent grounds for thinking that the dissenting thinkers’ beliefs are less likely to be correct than yours. If the other party to the disagreement is less rational than you are, or less well informed about than you are, then perhaps there will be such independent grounds for thinking that they are less likely to be correct than you are. However, if a moral evil demon has been at work, then the disagreement is explained simply by the fact that the two parties to the disagreement have different fundamental intuitions. In this case, there will be no such independent grounds for thinking that the other thinker is less reliable than you are. So

according to Sidgwick's principle, you should suspend judgment about these moral propositions.

In this way, we can give a good interpretation of the problem of the moral evil demons if we view it as resting on something like Sidgwick's principle. For this reason, I shall respond to the problem of the moral evil demons by arguing that Sidgwick's principle is in fact incorrect, and that we need a different model to account for the epistemic significance of moral disagreement.

3. Philosophical discussions of disagreement

Sidgwick was particularly interested in moral disagreement. Some of the other philosophers whom I have cited, on the other hand, such as David Christensen and Adam Elga, are interested in disagreement more generally. In fact, the general question about disagreement has recently been discussed by quite a number of epistemologists.⁵

It seems intuitively clear that there is a considerable variety of cases in which one thinker learns that another thinker disagrees with them about some question. In some cases, for example, you should regard the other thinker as clearly more expert than you are about the question at issue, and you should unhesitatingly defer to them; you may also have many different sorts of reason for regarding the other thinkers as more expert than you in this way. In other cases, it is rational for you to regard the other thinker as clearly mistaken; once again, there are many different reasons why you should think this. Then there are also many intermediate cases, where you should give some credence to the other thinker's belief, by weakening your own level of confidence in your own opinion, and shifting your opinion towards theirs, without completely deferring to the other thinker's view.

Many recent discussions of disagreement focus on the special case of disagreements among *epistemic peers*.⁶ There are various ways in which one may define the notion of an "epistemic peer". One simple way would be by simply stipulating that your epistemic peers have *exactly the same evidence* as you have, and are *equally rational* (either in the sense that they are equally rational in the particular process of thinking that led them to their opinion about the particular question that is at issue, or perhaps just in the sense that they are generally speaking no less disposed to rational thinking than you are yourself).

There is a sense in which the problem of moral evil demons that I am focusing on here is similar, since this problem focuses on cases in which there is moral disagreement between thinkers who are *equally rational* (in the strong sense that they are equally rational in the thinking that led them to their dissenting opinion about the question at issue) and *equally well informed about the non-moral facts*. However, it is not obvious that this set of cases is exactly the same as the set of the moral disagreements between epistemic peers, since I have not described the cases that I am focusing on in terms of "evidence". The reason for this is simple. There are at least two factors that influence what moral beliefs it is rational for one to hold. The first factor consists of one's non-moral beliefs, while the second factor consists of one's moral intuitions. It is not clear whether we should say that it is only the first of these

⁵ For some important recent contributions to this debate, see Richard Feldman (2006), Brian Frances (forthcoming), Peter van Inwagen (1996), Thomas Kelly (2005), Keith Lehrer (1976), Philip Pettit (2006), Alvin Plantinga (2000), Gideon Rosen (2001), and Brian Weatherson (ms).

⁶ This term features particularly prominently in the work of Kelly (2005).

two factors, or both of these two factors, that count as one's "evidence" for one's moral beliefs. Rather than getting into the question of what is the appropriate interpretation of the term 'evidence', I have avoided using this term in formulating the problem of the moral evil demons.

There are other ways of understanding what it means to call someone one of your "epistemic peers" with respect to a given question. For example, we might understand your epistemic peers to include everyone whom it was antecedently rational for you, prior to learning about any disagreement that you might have with them, to regard as equally likely to be correct about the question as you are. Alternatively, we might understand your epistemic peers to be everyone with respect to whom it was antecedently rational for you to regard it just as likely, if you and they disagree about the question at issue, that they are correct and you are wrong as that you are correct and they are wrong—that is, it is rational for you to have the same conditional probability, given the supposition that you and they disagree about this question, for the proposition that they are right about this question as for the proposition that you are right.⁷

These two further ways of understanding what it is for someone to be your "epistemic peer" are importantly different from each other, as we shall see later on. In part because there are all these different ways of understanding what it is to be someone's "epistemic peer", I shall not make much use of this term here. Even without explicitly focusing on disagreements between epistemic peers, my arguments will be relevant to the debates that have explicitly focused on disagreements between epistemic peers. Many of the participants in those debates—including Elga, Christensen, and Feldman—have articulated principles (like Sidgwick's principle) that apply to *all* cases in which one learns that another thinker disagrees with one's belief about a given question, and my arguments will be immediately relevant to the evaluation of those principles.

4. Do the moral evil demons pose a *special* problem about moral disagreement?

Is the problem that I have identified a special problem for moral belief? This is a crucial question for our discussion. Several philosophers think that we are unusually resistant to adjusting our moral beliefs in response to learning that other moral thinkers disagree with us; as these philosophers put it, it seems to them that we are much more "intransigent" in the face of such moral disagreement than we are in the face of other kinds of disagreement.⁸ If this is right, then we must either concede that most people are irrationally overconfident in their moral beliefs, or else we must argue that there is something highly special and unusual about the epistemology of moral belief.

It seems doubtful to me whether the case of moral disagreement is a special case in this way. There are many other areas of thought in which there are disagreements that are just as profoundly entrenched as moral disagreements. For example, there are some *theological* disagreements that do not obviously seem to involve any irrationality on either side (for example, consider a disagreement between

⁷ Adam Elga (2007) works with this latter understanding of what it is for someone to be your "epistemic peer".

⁸ For this point, see especially Kalderon (2005, 8-36). My colleague Alison Hills is also developing an account of moral epistemology that is designed to explain this allegedly special sort of "intransigence" on the basis of some allegedly special features of moral thought.

an atheist who rejects all arguments for the existence of god, and a deist who accepts a version of the cosmological argument). Similarly, it is far from obvious that all *philosophical* disagreements must involve any irrationality on either side (for example, consider the disagreements between the various rival theories of the semantics of vague expressions in natural language, or the disagreements between different theories of how to understand the possibility that there might have been “additional” objects, which do not exist in the actual world). There also seem to be disagreements about some of the hard questions of history and social theory that do not obviously involve any irrationality on either side, or any error or ignorance about the relevant uncontroversial facts.

In at least some of these cases, the proponents of the various rival views are just as prone to stick to their guns and to refuse to adjust their opinions when they learn that another thinker disagrees with them as they are in cases of moral disagreements. Moreover, the subject-matter of each of these disagreements is not in any obvious way a moral question; in many cases, it is not even a normative question of any other kind either. So it does not seem that there is anything unique or special about our tendency to be intransigent in the face of moral disagreement: there are many other non-moral questions on which we refuse to adjust our opinions even when we learn that other thinkers dissent from our opinion. It seems *prima facie* more plausible that what we have here is a general phenomenon, not something that is peculiar to moral disagreements. For this reason, I shall assume that the solution to the problem of the moral evil demons will not depend on any special feature of the epistemology of moral belief, but will depend on some much more general considerations about the epistemology of disagreement instead.

What makes it possible for different thinkers to reach different conclusions about a question, even if neither of the two thinkers is being in any way irrational, or ignorant or misinformed about the uncontroversial facts that are relevant to the question? The answer seems to be that it is the same phenomenon that is sometimes identified by means of the slogan that “theory is underdetermined by the data”.⁹ If we identify the “data” with the uncontroversial facts that are relevant to answering a given question—that is, the facts that both sides of the dispute rationally take for granted—then there are many theoretical questions that are not decided by the data alone. One’s beliefs about these theoretical questions will also be influenced by some other aspects of one’s overall state of mind—either by one’s pre-existing beliefs, or by one’s dispositions to have intuitions or impressions about various questions (such as which views are more or less plausible than others), or the like. If it is rationally permissible for these other aspects of one’s overall state of mind to influence one’s attitude to such theoretical questions, then this may help to explain how two equally rational thinkers may arrive at different views of such questions.

This point brings out a basic feature of the concept of rationality. A sort of relativism is in a way obviously true of the concept of rationality. So long as we reject relativism about truth, then a proposition is either true or not true *simpliciter*, without relativization to anything else. But it is not true in the same way of every proposition *p* that it is either rational to believe *p* or not rational to believe *p simpliciter*. On the contrary, it may be rational for one thinker to believe *p* at one time without its being rational for the thinker to believe *p* at another time, and without its being rational for other thinkers to believe *p* at any time. In this way, the rationality of believing *p* is obviously relative to a thinker and a time.

⁹ Compare Wright (1992, 157-168).

Still, even if this point explains how it is possible for equally rational thinkers to arrive at different conclusions about such questions, this point does not yet explain how these two thinkers should respond once they *learn* about their disagreement. This is the topic that I shall be focusing on in the remainder of this paper.

5. A general epistemological framework

To make progress with evaluating Sidgwick's principle, we will need to see what the theoretical alternatives to it might be, and what implications these rival principles will have, in the context of our general epistemological framework. So we will need to make a number of assumptions about the general epistemological framework within which we are working.

One of the main assumptions that I shall make here is that a version of what epistemologists call "*internalism*" about rationality is true.¹⁰ Roughly, this is the view that rationality supervenes on the relevant thinker's internal mental states. (By speaking of a thinker's "internal" mental states, I mean to exclude the so-called "factive" mental states, such as *knowing that p*, which by their nature are mental states that can only have a fact or true proposition as their object.) What it is rational for a given thinker to believe at a given time depends purely on the facts about what internal mental states the thinker has at that time, and what mental processes she is going through at that time. To put it another way, in evaluating a belief as rational or irrational, we are not evaluating the belief on the basis of its relation to the external world; instead, we are evaluating the belief purely on the basis of its relation to the thinker's other mental states.

A second assumption that I shall make is that there are two kinds of epistemic rule or principle, which I shall call "special" epistemic principles and "general" epistemic principles respectively. Special principles are principles that specify the way in which it is rational to respond to some quite specific type of mental state. General principles, on the other hand, apply quite generally to all beliefs whatsoever. For example, special principles may include the following: (i) the principle that it is rational to take one's sensory experiences at face value (at least in the absence of any special reasons for doubting that one is perceiving properly in the circumstances); (ii) the principle that it is rational to take one's apparent memories at face value (at least in the absence of any special reasons for doubting that one is remembering properly in the circumstances); and (iii) the principle that it is rational to take one's moral intuitions at face value (in the absence of any special reason for doubting that one's moral intuitions are reliable about the relevant question). General principles might include principles of logical consistency, deductive and inductive coherence, and the like.

One framework that makes this distinction between special and general principles especially clear is a Bayesian framework. Within a Bayesian framework, the *general* epistemic principles are those that require (i) that one's degrees of belief should be probabilistically coherent (that is, that it must be possible to represent those degrees of belief by means of a probability function), and (ii) that when one acquires new evidence, one should update one's degrees of belief by means of Bayesian conditionalization. However, a Bayesian framework will also need to assume that there are some *special* principles as well, in order to explain what it is to acquire

¹⁰ For arguments in favour of this sort of "internalism", see Wedgwood (2002b and 2006).

“evidence” at all. For example, perhaps one version of this Bayesian framework will suppose that all evidence is acquired directly through sensory observation—but in that case, it will be committed to the existence of a special principle to the effect that it is rational to treat one’s sensory observations as evidence in this way.

For my purposes, however, the classical Bayesian framework is less natural than the less standard variant of the framework in which rational believers update their beliefs by Jeffrey conditionalization (instead of classical Bayesian conditionalization). The classical Bayesian framework requires a definite notion of “evidence”, and maintains that the beliefs that it is rational to hold are determined solely by one’s prior probabilities and one’s “evidence”. Moreover, if one is fully rational, then for every proposition p that forms part of one’s “evidence”, one would have to be *maximally* confident in p —as confident in p as one is in the simplest logical truths. By contrast, the variant of the Bayesian framework that invokes Jeffrey conditionalization allows that certain events that may not themselves count as one’s acquiring any “evidence” (such as the event of one’s having a sensory experience, or a memory, or a moral intuition) can change the degree to which it is rational to believe a proposition p , without making either p or $\neg p$ as completely certain as a logical truth. Then this approach implies that to maintain coherence through one’s whole system of beliefs, one should revise one’s degrees of belief in all the other relevant propositions in accordance with Jeffrey conditionalization.¹¹

In this way, the variant of the Bayesian framework according to which the rational way of updating one’s beliefs is by means of Jeffrey conditionalization has no need to appeal to any notion of “evidence” at all. All that it requires is that some event that does not consist in one’s acquiring evidence can change the degree to which it is rational for one to believe a given proposition; the knock-on effects of this change for the rest of one’s beliefs are then explained by Jeffrey conditionalization. (According to an *internalist* version of this approach, this event that changes the degree to which it is rational to believe this proposition will always be some *internal* mental event—such as an experience or a memory or an intuition or the like.) As I explained earlier in Section 3, I have avoided speaking of “evidence” here; so it would be more natural for me to opt for the version of the Bayesian approach that involves Jeffrey conditionalization than to go for the classical Bayesian approach.

6. The epistemic significance of information about others’ beliefs

Within the context of the general framework that I have just outlined, Sidgwick’s principle clearly counts as a *special* epistemic principle. In effect, Sidgwick’s principle gives a *special* significance to information about others’ beliefs. That is, Sidgwick’s principle tells one how one should respond to a specific sort of mental state—namely, to the mental state of (rationally) believing that another thinker believes p (where one had previously believed $\neg p$ oneself). It is not a general principle that applies quite generally to all of one’s beliefs whatsoever.

Admittedly, Sidgwick’s principle is not a special principle about when it is rational to *form* a new belief (like the principle that it is rational to take one’s sensory experiences at face value, at least in the absence of special reasons for doubt). Sidgwick’s principle is concerned with when we are rationally required to *abandon* a belief, or in other words a principle about when our past beliefs are *defeated*:

¹¹ For an account of Jeffrey conditionalization, see Jeffrey (1983, chap. 11).

according to Sidgwick's principle, the information that another person believes p invariably defeats your prior belief in any proposition that is incompatible with p , unless there is *independent* reason for you to believe that the other thinker is less likely to be right about the question than you are yourself. Still, this is a principle that gives a special epistemic significance to information about the beliefs of others.¹²

What rationale there could be for this special principle? This is a particularly pressing problem for the proponents of Sidgwick's principle, since there seems to be a much simpler way to conceive of the epistemic significance of information about others' beliefs—specifically, one could conceive of such information as simply one more piece of empirical information like any other, the epistemic significance of which is explained purely by *general* epistemic principles instead. On the face of it, there seems to be significantly more to be said in favour of this rival approach than in favour of the approach that is based on Sidgwick's principle.

Consider for example how it seems rational to respond to information about the state of measuring instruments, such as pieces of litmus paper. It would surely be misguided to postulate any special epistemic principles that are concerned solely with information about the states of pieces of litmus paper. Instead, we can explain how it is rational for you to respond to this information by appealing purely to general principles instead. If the relevant general principle is a version of conditionalization (such as Jeffrey conditionalization), then the rational way for you to respond to information that the piece of litmus paper has turned red is determined by the *conditional beliefs* that it was antecedently rational for you to have—such as the conditional beliefs that it was antecedently rational for you to have in the various relevant propositions, on the supposition that the litmus paper would turn red, and so on. Given an internalist approach, the conditional beliefs that it is antecedently rational for you to hold themselves reflect all the internal mental factors that have had an influence on what it is rational for you to believe—your sensory experiences, your apparent memories, your background beliefs, and so on. For example, if your mental life has been anything like mine, these internal mental factors have made it rational for you to have a high conditional degree of belief that the liquid into which the litmus paper was inserted is an acid, given the supposition that the litmus paper turned red.

In this way, my past mental life has made it rational for me to have a large stock of conditional beliefs about the world and how it works (including the dispositions of litmus paper). These conditional beliefs presumably include conditional beliefs about other people and how their minds work, and in particular about the circumstances in which people's beliefs are reliable, and in which they are unreliable. For example, the conditional beliefs that it is rational for me to have tell me that people's sensory perceptions are usually fairly reliable, whereas unless they have a special expertise, their beliefs about abstruse theoretical matters (such as about the age of the universe, or the correctness of various philosophical interpretations of quantum mechanics) are usually much less reliable. Given that my past mental life has made it rational to have this large stock of prior conditional beliefs, it could be that the way in which it is rational for me to respond to information about the beliefs of others is completely determined by these conditional beliefs, in accordance with some completely general epistemic principles (such as Jeffrey conditionalization).

¹² This issue, about whether there is a "special" principle defining the epistemic significance of information about other people's beliefs, or whether the significance of this information can be completely explained by general epistemic principles, clearly mirrors the debate between "reductionists" and "anti-reductionists" in the epistemology of *testimony*; for this debate, see especially Fricker (1995).

There are several advantages to this approach, according to which the rational response to information about the beliefs of others is determined purely by general epistemic principles in this way. First, this approach is in an obvious way more economical: if Sidgwick's principle is true, it would surely call for some explanation or rationale; since this approach dispenses with any such special principle, it postulates fewer phenomena that cry out for explanation. Some special principles can be fairly easily explained. In particular, there must be some special principles specifying the epistemic significance of *non-doxastic* states (like sensory experiences, apparent memories, intuitions, and the like): the epistemic significance of non-doxastic states cannot be captured by general epistemic principles, and can only be accounted for by special principles. But clearly this explanation does not apply in the case of one's beliefs about the beliefs of others. The epistemic significance of one's beliefs about the beliefs of others can easily be captured by general principles; there seems no obvious need for any special principle here.

Secondly, this approach can give an illuminating explanation of the wide variety of ways in which we respond to learning about the beliefs of others. In some cases, it was antecedently rational for you to believe that whatever belief the other thinker has on the question, it is far more likely that the other thinker is right about the question than that you are right. (Perhaps this is because the question is about the facts of that other thinker's personal life, or about some topic on which that other thinker has world-renowned expertise.) In these cases, you should simply defer immediately to the other thinker's view. In other cases, the situation is reversed: it was antecedently rational for you to believe that whatever belief the other thinker has about the question, you are much more likely to be right than they are. (Perhaps this is because the disputed question concerns the grammar of a language of which you are a native speaker, while the other thinker has only been studying the language for a few months.) Then of course there is a wide spectrum of intermediate cases, where your prior rational conditional beliefs make it rational for you to respond to the information that the other thinker disagrees with you by *weakening* your degree of belief on the disputed question, but without simply deferring to the other thinker's view. In general, which of these ways of responding to the information that the other thinker disagrees with you is rational in the circumstances is simply determined by the conditional beliefs that it was antecedently rational for you to have; and that in turn is determined by the totality of your mental states (including your background beliefs, your experiences, your memories, and so on). There is no limit in principle to the ways in which the totality of your mental states may have determined which conditional beliefs were antecedently rational for you; in particular, both general epistemic principles and special epistemic principles may be involved in explaining how these mental states determined which conditional beliefs it was rational for you to have.

It seems to me that there is precisely the same spectrum of cases when the disagreement concerns a *moral* question as when it concerns any other question. In some cases, your prior rational conditional beliefs will lead you to regard the other thinker's moral sensibility as vicious and corrupt. In other cases, one's prior rational conditional beliefs will lead one to regard the other thinker as more reliable about moral questions of the relevant kind than one is oneself. For example, suppose that you know that you and the other thinker agree about almost all moral questions of this kind, but that in the few cases in which you have initially disagreed, you have always in the end been persuaded that she was right and you were in fact wrong. Then it will

presumably be rational for you to treat her intuitions as more reliable than your own.¹³ Between these two extremes, there are many intermediate cases where rationality will require you to *weaken* your confidence about your moral opinion without requiring that you simply defer to the other thinker.

Thirdly, this approach it can also explain why it is rational in some cases to be intransigent in the face of disagreement. This approach does *not* require that your reason for thinking the other thinker to be less reliable than you are yourself must be *independent* of the reasoning that led you to your view on the disputed question. In some cases, even though you might initially have thought it highly likely that the other thinker would be right about the question at issue (perhaps because in your experience so far, the other thinker has always seemed impressively intelligent and well informed), the very fact that the other thinker believes p may rationally convince you that the other thinker is less reliable than you had previously thought. In these cases, you attach a high *unconditional* probability to the hypothesis that the other thinker will be right about the question—but this is only because you are confident that the other thinker will believe $\neg p$, which you regard as most probably the right answer to the question. You do *not* attach a high *conditional* probability to the proposition that the other thinker is right, on the supposition that the other thinker believes p (a belief that you think most probably wrong). Even if initially—before you find out that I disagree with you—you rationally thought that I was just as likely to be right as you, the information that I believe p may *by itself* give you sufficient reason to think that I am probably less reliable than you are.

This third sort of case shows how important it is to be clear about the definition of what it is for someone to count as one of your “epistemic peers”. On the one hand, suppose that we say that for you to regard another thinker as your “epistemic peer” (with respect to a given question) is for you to attach an equally high *unconditional* probability to the hypothesis that that thinker will be right about that question as to the hypothesis that you will be right about that question. Then, even if you start out by rationally regarding me as your epistemic peer with respect to a given question, it may be quite rational for you to respond to the information that I believe p (which you believe clearly false) by ceasing to regard me as your epistemic peer with respect to that question.

On the other hand, suppose that we say that for you to regard me as your “epistemic peer” with respect to a given question is for you to assign exactly the same *conditional* probability, on the supposition that you and I disagree about the question, to the proposition that I am right as to the proposition that you are right. Suppose that we also assume that the rational way to respond to new information is by some form of conditionalization. Then if you rationally regard me as your epistemic peer in this sense, it could *not* be rational for you to respond to the information that you and I disagree by concluding that you are more likely to be correct than I am. This point is not in any way a qualification of this way of understanding the epistemic significance of information about other thinkers’ beliefs. This interpretation of what it is for you to regard someone as your “epistemic peer” makes it very unlikely that you will regard many people as your epistemic peers. On any less demanding interpretation of what it is for me to be your epistemic peer, it may be quite rational in certain cases for you to downgrade your assessment of my epistemic standing in relation to your own precisely in response to the information that I disagree with you.

Indeed, in some cases, it may be rational for *both* sides in a disagreement to

¹³ Thus, I need not disagree with the central claims of Karen Jones (1999).

regard it as more likely that the *other* side is mistaken about the disputed question than that they are mistaken themselves. It may be that some of the cases of apparently irresolvable disagreement that we considered earlier, in Section 1, are cases of this kind. It may be part of the explanation of people's intransigence in the face of these disagreements that such intransigence is in fact a rational response for all the parties involved. Since many people do respond to disagreements with this sort of intransigence, without believing that there is anything irrational about their response, the approach that I am describing here seems significantly less revisionary than the rival approach that is based on Sidgwick's principle. The fact that this approach is so much less revisionary than its rival seems to me another attractive feature of this approach.

7. Epistemic egoism?

It may seem that the proponent of Sidgwick's principle has an easy way of responding to the complaint that there must be some explanation or underlying rationale for this special epistemic principle. Indeed, the explanation has already been sketched, at least in outline, by Allan Gibbard (1990, 176–181). The rough idea behind Gibbard's account is straightforward. First, Gibbard argues that it is impossible to get by as a thinker without having a kind of "self-trust"—that is, trust in one's own intellectual capacities and dispositions. Then it is argued that this sort of "self-trust" will commit one to a general "fundamental trust" of all minds, including other minds as well as one's own. This might seem to support Sidgwick's principle, because if we are committed to this sort of fundamental trust in all minds, then it seems that one must always attach some credence to the beliefs of other thinkers, unless one has some independent reason for discounting those beliefs as unreliable.

According to Gibbard, we can distinguish between two kinds of sources of information. On the one hand, there are those sources of information in which it is rational to have a sort of "fundamental trust"; that is, it is rational for us simply to be disposed to believe the pieces of information that derive from those sources, even in the absence of any *independent* reasons for regarding those sources as reliable. On the other hand, there are those sources of information that we "trust" (that is, we are disposed to believe the pieces of information that derive from those sources), but only because we have such independent reasons for regarding these sources as reliable.

It is clear that we have a fundamental need for a sufficient stock of beliefs about the world in the light of which to live and act. But we could never have any beliefs at all unless we placed such "fundamental trust" in at least *some* sources of information. Gibbard (1990, 178–9) argues that we need to place such fundamental trust quite generally in all of one's "judgments": that is, even in the absence of any independent reason in favour of regarding one's beliefs as reliable, one should continue to rely on one's beliefs as though they really were reliable—at least so long as one does not have any special defeating reasons for thinking that one's beliefs are not reliable in the circumstances.

Having argued for this sort of "self-trust", Gibbard (1990, 179–81) goes on to argue that such "self-trust" commits one to a general fundamental trust in all minds as such. There are two main links that connect self-trust with a more general trust in minds as such. First, it may seem that fundamentally the only rational ground for trusting one's own beliefs is simply that they are beliefs: there is surely nothing special about the fact that these beliefs are one's own. Secondly, since one has

acquired such an enormous number of one's beliefs from what one was told by other people while one was growing up, if one were to give up this fundamental trust in the beliefs of other people, one would be faced with the practically impossible task of reconstructing one's whole belief system without relying on any of the beliefs that one acquired through one's earlier trust in others. So it may seem that we have no real alternative to having a fundamental trust in all beliefs as such, including the beliefs of others.

One point where this argument could be resisted is by pointing out that it is not clear that we really need to have this sort of fundamental trust in *all* of one's own beliefs as such. Perhaps one only needs to have this sort of fundamental trust in a certain *subset* of one's beliefs (such as certain pivotal entrenched background beliefs, perhaps), and also in certain *non-doxastic* mental states, such as one's sensory experiences, one's apparent memories, one's moral intuitions, and so on. If one is not committed to any such fundamental trust in all of one's beliefs as such, then there will be no reason to think that one is also committed to any such fundamental trust in all the beliefs of others.

Still, even if Gibbard's argument is mistaken in this way, we could still argue for a variant of Sidgwick's principle. Let us assume that it is true that one is inevitably committed to having this sort of fundamental trust in one's own moral intuitions. Then it may seem that one will also be committed to having a comparable trust in *all* moral intuitions, including other people's moral intuitions.

It is true that I have no reason to think that the mere fact that some moral intuitions are *mine* makes those intuitions any more reliable than anyone else's. But it does not follow that if it is rational for me to have this sort of fundamental trust in my own moral intuitions, it must also be rational for me to have the same sort of fundamental trust in everyone else's intuitions. If it is rational for me to have this sort of fundamental trust in my own current moral intuitions, there must indeed be some feature of these intuitions that explains why it is rational for me to trust them in this way. But the feature of these intuitions that explains this need not consist simply in their being moral intuitions that are had by someone at some time. Another part of what explains why my current intuitions have the special rational role for me that they have is that it is possible for me to base my current formation of a moral belief *directly* on these intuitions.

It does not seem possible for me currently to form a moral belief *directly* on the basis of *your* moral intuitions. At best, I can only directly base my current formation of a moral belief on my *beliefs* about your moral intuitions. On the other hand, it *is* possible for me currently to form a moral belief directly on the basis of *my own current* moral intuitions. Moreover, it seems that we are disposed to be guided by our moral intuitions towards forming the corresponding moral beliefs: if I currently have a moral intuition, that moral intuition will immediately incline me to accept the corresponding moral belief (unless I have some special reason for doubting that intuition). On the other hand, there is no such immediate tendency for *your* moral intuitions to incline me to accept the corresponding moral beliefs; even my own *beliefs* about your moral intuitions do not seem immediately to incline me to accept the corresponding moral beliefs.

At least assuming what epistemologists call an "internalist" view of rationality, the facts that make it rational for one to revise one's belief in a certain way must be capable of *directly guiding* one towards revising one's beliefs in that way. But as I have argued, the fact that someone else has a certain mental state cannot directly guide one in one's revisions of one's beliefs. It is only one's own mental

states that can do this. So it is simply out of the question that other people's intuitions should play the same role in rationally guiding my reasoning as my own intuitions. At most, it might be that *my beliefs* about other people's intuitions should play the same role in guiding my reasoning as my own intuitions. But *my intuitions* seem to be such different mental states from *my beliefs about other people's intuitions* that it is implausible to claim that they should play exactly the same role in guiding my reasoning.

Indeed, it is striking that in these respects, it is only my own *current* intuitions that can play this role in guiding my current reasoning. I cannot directly base the formation of a new moral belief on my *past* intuitions (at best I can directly base the formation of a new moral belief on my *memory* of those past intuitions). Similarly, I cannot base my beliefs directly on my *future* intuitions, but only on my *expectation* of future intuitions. Moreover, if I either remember having a past intuition or expect to have a future intuition, but do not have the intuition itself now, then this memory or expectation will not now immediately incline me to accept the corresponding belief.

It might seem strange to claim that it is not rational to have the same sort of "fundamental trust" in our past moral intuitions as in our present intuitions. But the metaphor of "trust" is misleading here. What this claim really amounts to is the claim that whereas it can be rational to form moral beliefs directly on the basis of one's current intuitions—even without any additional independent reason for regarding those intuitions as reliable—it cannot be rational to form a moral belief directly on the basis of one's memory of a past intuition unless one has some further reason for regarding that past intuition as reliable. Moreover, we can certainly admit that it is rational to be guided by one's *entrenched current background beliefs*; and no doubt some of these entrenched background beliefs will themselves reflect one's past moral intuitions.

With regard to our future intuitions, it does not seem so strange to me that even if one receives the information that one will have a certain intuition in the future, one's response to this information should be guided by one's rational assessment of whether one's intuitions can be expected to become more or less reliable in future. There is also no need for this assessment to be independent of one's current intuitions. Indeed some cases, the very information that I will come to have a certain intuition will itself give me reason to think that my moral sensibilities will deteriorate. For example, suppose that I receive the information that in the future, I will have the intuition that it is an admirable form of tough-mindedness for the police to have a policy of torturing those whom they suspect of serious crimes. This information would seem to me all by itself to make it rational for me to think that my moral intuitions will probably be less reliable in future than they now are.

Gibbard's second argument is closely akin to a well-known argument that is given by C. A. J. Coady (1992) in the course of his argument for a version of the principle of *credulity*—that is, roughly, the principle that it is rational for us to believe everything that we are told unless we have some special positive reason to suspect that our informant is unreliable.¹⁴ Coady points out that we only came to master a language by coming to believe most of what we were told while we were growing up, and that we acquired many of these childhood beliefs in a fundamentally uncritical way. It may seem then that unless the principle of credulity is true, we will be rationally required to pursue the Cartesian project of reconstructing our whole belief-

¹⁴ The inspiration for this principle of credulity derives at least in part from Thomas Reid (1764, chapter 6, esp. section 24).

system from the ground up, without any initial reliance on anything that we have learnt from others. Since that project does not seem feasible, it seems that we cannot really be rationally required to pursue it, and so—according to this argument—the principle of credulity must in fact be sound.

As the literature on the epistemology of testimony has made clear, there is a straightforward response that is available to those who are sceptical of the principle of credulity. Perhaps it is true that young children have to pass through a phase of uncritically believing everything that they are told if they are to learn a language and acquire a sufficiently rich understanding of the world. But it does not follow that once these children have acquired the status of being fully rational agents, they should continue to be so uncritical. Perhaps we can't help but continue to rely on many of the beliefs that we acquired uncritically while we were a child. But it doesn't follow that we should continue to acquire beliefs in such an uncritical way.

What makes it possible for this response to Gibbard's second argument to be coherent is that there is at least some plausibility in the suggestion that epistemology should treat our *continuing to rely on our old background beliefs* differently from our *currently forming new beliefs*. Since we are not engaged in Descartes' project of "pure inquiry", we are not required to reconstruct all our background beliefs from the ground up. We may continue to rely on our entrenched background beliefs even if we cannot reconstruct them all in this way. But it does not follow that all the ways in which we originally acquired those old background beliefs are ways in which it is rational for us now, as mature adult thinkers, to form new beliefs about the world. The fact that we absorbed so many beliefs from other people while we were children shows at most that it must be rational for us to continue relying on the specific beliefs that we absorbed from others, not that we are rationally required to have a fundamental trust in all minds as such. One may have had to pass through a period of childish credulity while growing up, but as an adult thinker one should be able to put this sort of credulity aside.

In earlier work (Wedgwood 2007, chap. 11), I described my position on the epistemic significance of disagreement as a view that it is rational to have a sort of *egocentric bias* in forming beliefs. In a sense, this is an accurate description of my position. It is rational for my background beliefs, experiences, memories and intuitions to guide me *directly* in a way in which it is simply not possible for your beliefs, experiences, or intuitions to guide me directly. That is, my thinking can be guided by my mental states even if I am not thinking *about* those mental states at all. Indeed, it is *inevitable* that if I form any beliefs at all, my thinking will be at least sometimes directly guided by my own current beliefs, experiences, memories or intuitions in this way.

Moreover, in my view, it is rational for me, in being directly guided by my own current beliefs, experiences, memories and intuitions, to have a tendency to form new beliefs, endorsing the content of those states, directly on the basis of those states, even in the absence of any independent reason for regarding those states as reliable guides to the truth. We could metaphorically describe this tendency as a kind of "fundamental trust" in those states (although it should be emphasized that this "trust" in one's mental states need not involve *thinking about* one's mental states at all—one's attention may be fixed on the world, not on one's own mind). On the other hand, in rejecting the idea of any "special" epistemic principles defining the epistemic significance of information about the beliefs of others, I deny that it is rational to have the same sort of "fundamental trust" in the mental states of other people. So my view

does give a profoundly different significance to one's own current mental states compared to the mental states of other people.

Unfortunately, my description of my view may have encouraged some misunderstandings. It is *not* my view that your beliefs and my beliefs both function for you as reasons or evidence of fundamentally the same kind, but it is rational for you to give greater weight to your beliefs than to mine. On the contrary, my view is that the role of your current beliefs and intuitions in guiding your thinking is profoundly different from the role of my beliefs or intuitions (and indeed the role of your own past or future beliefs or intuitions as well). The image of different "weights" for two bodies of evidence of essentially the same kind is in fact a travesty of my view.

Thus, I am also not saying, absurdly, that when you learn that you and I disagree about something, it is rational for you to think to yourself: 'I'm me, and he's not; so I'm probably right, and he's probably wrong.' On my view, when you consciously entertain a belief or an intuition with the content *p*, your attention is not on the fact that you are entertaining this belief or intuition, but rather on *p* itself. You might express this intuition or belief by saying something like 'At least probably, *p*'. So in entertaining this intuition or belief, your attention is on the facts, as they appear to you probably to be. According to my picture, when you learn that you and I disagree about *p*, it may in some cases (although certainly not in all cases) be rational for to continue relying on your original belief. So, in these cases, it may be rational for you to think to yourself: 'At least probably, *p*; but he believes that it is *not* the case that *p*: so he's probably wrong.' That does not seem to me to be in any way an absurd or irrational response to this information.

The conclusion of this essay, then, is this. It seems plausible that there is a sort of rational asymmetry between one's own moral intuitions and the intuitions of other people: it is rational to have a special sort of "fundamental trust" in one's own intuitions, but it is not even possible to have the same sort of "trust" in the intuitions of others. In consequence, even though we know full well that there is widespread disagreement about fundamental moral issues, and that this shows that in all likelihood "moral evil demons" have been at work, this knowledge may not always require us to suspend judgment about these moral issues completely. It may indeed often require us to *weaken* our degree of confidence in our beliefs about those moral issues. But at least sometimes, it may be rational for each of us to continue having more confidence in the propositions that we believe than in the incompatible propositions that are believed by those who disagree with us.

Of course, if relativism is false, then whenever two thinkers disagree, at least one of them has as a matter of fact got things wrong, and believes something false. But it is perfectly possible for false beliefs to be rational. Indeed, when you and I disagree, we could even recognize that each of us is rational—while I think that my belief is rational and true and your belief is rational but false, and you think that your belief is rational and true and my belief is rational but false. Even if we do not respond to learning about disagreement by abandoning our original belief, we do not have to be so dogmatic as to conclude that the other thinker is irrational—although we are committed to thinking that the other thinker is mistaken. Since this account of the epistemic significance of moral disagreement seems to be reasonably coherent and plausible, we may conclude that it can solve the problem with which we began. Even given the deep and irresolvable nature of some moral disagreements, the metaethical views that reject relativism can avoid being committed to any sweeping form of moral

scepticism.¹⁵

¹⁵ Earlier versions of this paper were presented to a Metaethics Workshop at the University of St Andrews and to the Fourth Annual Metaethics Workshop at the University of Wisconsin in Madison. I am grateful to the members of those audiences, and also to my colleagues David Charles, Bill Child, Antony Eagle, Lizzie Fricker, John Hawthorne, and Timothy Williamson, for helpful comments.

References

- Boyd, Richard (1988). "How to Be a Moral Realist", in Geoffrey Sayre-McCord, ed., *Essays on Moral Realism* (Ithaca, New York: Cornell University Press).
- David Christensen (2007) "Epistemology of Disagreement: the Good News", *Philosophical Review* 116 (2007): 187—217.
- Coady, C. A. J. (1992). *Testimony: A Philosophical Study* (Oxford: Clarendon Press).
- Elga, Adam (forthcoming). "Reflection and Disagreement", forthcoming in *Noûs*.
- Feldman, Richard (2006). "Epistemological Puzzles about Disagreement", in Stephen Hetherington, ed., *Epistemology Futures* (Oxford: Clarendon Press).
- Frances, Bryan (forthcoming). "Live Skeptical Hypotheses", forthcoming in John Greco, ed., *The Oxford Handbook of Skepticism* (Oxford: Clarendon Press).
- Fricker, Elizabeth (1995). "Telling and Trusting: Reductionism and Anti-Reductionism in the Epistemology of Testimony", *Mind* 104: 393-411.
- Gibbard, Allan (1990). *Wise Choices, Apt Feelings* (Oxford: Clarendon Press)
- Harman, Gilbert, and Thomson, Judith Jarvis (1995). *Moral Relativism and Moral Objectivity* (Oxford: Blackwell).
- Jeffrey, Richard (1983). *The Logic of Decision*, 2nd edition (Chicago: University of Chicago Press).
- Jones, Karen (1999). "Second-Hand Moral Knowledge", *Journal of Philosophy* 96 (no. 2, February): 55-78.
- Kalderon, Mark (2005). *Moral Fictionalism* (Oxford: Clarendon Press)
- Kelly, Thomas (2005). "The epistemic significance of disagreement", in John Hawthorne and Tamar Gendler, eds., *Oxford Studies in Epistemology*, vol. 1 (Oxford: Oxford University Press).
- Lehrer, Keith (1976). "When Rational Disagreement is Impossible", *Noûs* 10 (1976), 327-332.
- Miller, Richard W. (1992). *Moral Differences* (Princeton, New Jersey: Princeton University Press).
- Pettit, Philip (2006). "When to defer to majority testimony—and when not", *Analysis* 66 (2006), 179-187.
- Plantinga, Alvin (2000). "Pluralism: A defense of religious exclusivism", in Philip L. Quinn and Kevin Meeker, eds., *The Philosophical Challenge of Religious Diversity* (Oxford: Oxford University Press), 172-192.

- Reid, Thomas (1764). *An Inquiry into the Human Mind on the Principles of Common Sense* (Edinburgh: Kincaid and Bell).
- Rosen, Gideon (2001). "Nominalism, naturalism, epistemic relativism", *Philosophical Perspectives* 15: 69-91.
- Sidgwick, Henry (1907). *The Methods of Ethics*, 7th edition (London: Macmillan).
- (2000). *Essays on Ethics and Method*, ed. Marcus G. Singer (Oxford: Clarendon Press).
- van Inwagen, Peter (1996). "It is wrong, everywhere, always, and for anyone, to believe anything upon insufficient evidence", in Jeff Jordan and Daniel Howard-Snyder, eds., *Faith, Freedom and Rationality: Philosophy of Religion Today* (London: Rowman & Littlefield), 137-153.
- Weatherson, Brian (ms). "Disagreeing about Disagreement".
- Wedgwood, Ralph (2002a). "The Aim of Belief", *Philosophical Perspectives* 16: (2002), 267-297
- (2002b). "Internalism Explained", *Philosophy and Phenomenological Research* 65 (2002), 349-369
- (2006). "The Internal and External Components of Cognition", in Robert Stainton, ed., *Contemporary Debates in Cognitive Science* (Oxford: Blackwell), 307-325.
- (2007). *The Nature of Normativity* (Oxford: Clarendon Press).
- Williamson, Timothy (2000). *Knowledge and its Limits* (Oxford: Clarendon Press).
- Wright, Crispin (1992). *Truth and Objectivity* (Cambridge, Massachusetts: Harvard University Press).