

- Timmons, M. 1999. *Morality without foundations*. New York: Oxford University Press.
- Wiggins, D. 1988. Truth, invention, and the meaning of life. In *Essays on moral realism*, ed. G. Sayre-McCord, 127–165. Ithaca: Cornell University Press.
- Wright, C. 1992. *Truth and objectivity*. Cambridge, MA: Harvard University Press.

Beyond the Error Theory

Michael Smith

In *Ethics: Inventing Right and Wrong*, published just over 30 years ago, John Mackie famously argued for the error theory (Mackie 1977). Though the argument initially met with considerable skepticism (see, e.g., Blackburn 1993; McDowell 1985), in the years that followed many theorists came to the conclusion that Mackie had things more or less right (see e.g., Lewis 1989; Garner 1994; Joyce 2001). But which of these views is correct? Should we all be error theorists? Or is the renewed admiration for Mackie's argument itself mistaken? To anticipate, my somewhat tentative suggestion will be that Mackie's argument fails. Since I have been tempted to believe the error theory myself in the past (Smith 2006), this represents something of a shift in my own thinking.

1 An Outline of Mackie's Argument

Mackie's argument for the error theory proceeds in two stages. He begins by pointing out that our concept of a moral value is the concept of a feature of things that is at once both objective and prescriptive. But, he then goes on to argue, general reflections of a metaphysical and epistemological kind show that nothing has such a feature: The concept of an objective and prescriptive feature isn't instantiated. As we will see, the reasons he offers make it seem that the conclusion would have to be necessary, so the upshot, if the argument works, is not just that nothing has moral value, but that nothing could have such value.

As even this briefest of outlines makes clear, the real power of Mackie's argument is that it is addressed to people who are antecedently engaged in ordinary moral thought and talk in blissful ignorance of the error that it is the aim of his argument to lay bare. His strategy is to get those people to agree first to the conceptual claim – this fixes what it is that they are thinking and talking about – and then to the substantive reasons he offers for supposing that there could be nothing like that.

M. Smith (✉)
Department of Philosophy, Princeton University, Princeton, NJ, USA

Mackie's argument is thus an internal critique of morality: It purports to show that someone who is committed to morality shouldn't be so committed, and it purports to show this on terms that that person can himself recognize. Let's consider the two stages of the argument in more detail.

Mackie's defense of the conceptual claim consists in a rehearsal of what he sees as "the main tradition of European moral philosophy from Plato onwards," a tradition which, he tells us,

... has combined the view that moral values are objective with the recognition that moral judgments are partly prescriptive or directive or action-guiding. Values themselves have been seen as at once prescriptive and objective. In Plato's theory the Forms, and in particular the Form of the Good, are eternal, extra-mental, realities. They are a very central structural element in the fabric of the world. But it is held also that just knowing them or "seeing" them will not merely tell men what to do but will ensure that they do it, overruling any contrary inclinations. ... Similarly, Kant believes that pure reason can itself be practical, though he does not pretend to be able to explain how it can be. Again, Sidgwick argues that if there is to be a science of ethics—and he assumes that there can be, indeed he defines ethics as "the science of conduct"—what ought to be "must in another sense have objective existence: it must be an object of knowledge and as such the same for all minds"; but he says that the affirmations of this science "are also precepts," and he speaks of happiness itself as "an end absolutely prescribed by reason." (Mackie 1977, pp. 23–24)

Though Mackie seems to think that these diverse formulations are all in some sense equivalent, in what follows I will focus on just one of them, namely, Sidgwick's.

In Sidgwick's terms, Mackie's claim that our concept of moral value is the concept of an objectively prescriptive feature of things amounts to the claim that to conceive of (say) happiness as a moral value is to conceive of happiness itself as having the feature of *being an end that is absolutely prescribed by reason*.

Mackie's argument from this conceptual claim about moral value to the conclusion that there is no such value is brevity itself. He dubs it "the argument from queerness" and he tells us that it

... has two parts, one metaphysical, the other epistemological. If there were objective values, then they would be entities or qualities or relations of a very strange sort, utterly different from anything else in the universe. Correspondingly, if we were aware of them, it would have to be by some special faculty of moral perception or intuition, utterly different from our ordinary ways of knowing everything else. These points were recognized by Moore, who spoke of non-natural qualities, and by the intuitionists in their talk about a "faculty of moral intuition." Intuitionism has long been out of favour, and it is indeed easy to point out its implausibilities. What is not so often stressed, but is more important, is that the central thesis of intuitionism is one to which any objectivist view of values is in the end committed: Intuitionism merely makes unpalatably plain what other forms of objectivism wrap up. Of course the suggestion that moral judgements are made or moral problems solved by just sitting down and having an ethical intuition is a travesty of actual moral thinking. But, however complex the real process, it will require (if it is to yield authoritatively prescriptive conclusions) some input of this distinctive sort, either premises or forms of argument or both. (Mackie 1977, p. 38)

Though Mackie goes on to consider various replies to this argument, he doesn't really say anything more to spell it out beyond what he says in this passage. So what exactly is the argument?

Let's begin by restating Mackie's argument in Sidgwickian terms. Though we believe that happiness has the feature of being an end that is absolutely prescribed by reason, Mackie seems to be saying, the idea that happiness is such an end is both metaphysically and epistemologically queer. Why is it metaphysically queer? Mackie doesn't explicitly say, but we can imagine what he is thinking. Ends are the sorts of things that each of us has, in so far as we aim at, or desire, different things. But while it is therefore true that some of us have happiness as an end, as some of us do desire happiness, the claim that happiness has the feature of being an end absolutely prescribed by reason, "the same for all minds," would have to be made true by some further fact about happiness, a fact beyond this purely descriptive psychological fact. Yet what further fact is there?

At this point, Mackie seems to just throw up his hands. He cannot see what further fact there could be. Or, more accurately, he cannot think of anything beyond its being a brute further fact, a Moorean non-natural fact the like of which he can make no sense (Moore 1903). This seems to be the metaphysical queerness he has in mind. Moreover, since he can make no sense of what kind of fact would make true the claim that happiness is an end absolutely prescribed by reason, he cannot think of any way in which we might come to know about such a further fact either. Or, more accurately, he cannot think of anything beyond our possessing a special faculty which enables us to detect non-natural features, a faculty the like of which he can make no sense either. This seems to be the epistemological queerness that he has in mind.

I said earlier that the reasons Mackie gives for thinking that there are no moral values would, if they were successful, make the conclusion necessary. We can now see why this is so. The problem with Moorean non-natural qualities isn't that there aren't any such things as a matter of contingent fact. The problem is that we can literally make no sense of them: There are no possible worlds in which objects have such qualities (though see Shafer-Landau 2003). Since Mackie appears to think that the existence of objectively prescriptive features requires the existence of such non-natural qualities, it follows, if he is right, that there could be no objectively prescriptive features either.

To sum up: Mackie's argument from queerness consists in an analysis of the concept of moral value together with a pair of challenges which purport to show that we can make no real sense of how that concept could be instantiated. We are therefore left with the conclusion that there is nothing beyond facts like the purely descriptive psychological fact that some of us do indeed have happiness as an end. Suppose we were antecedently committed to morality. Mackie's argument purports to show that we shouldn't be. Moreover, it purports to show this on terms that we can ourselves recognize. If we are to respond to his argument then we must therefore either take issue with his analysis of value or else provide answers to the metaphysical and epistemological challenges.

2 Value or Obligation?

Let's begin with an initial worry about the analysis. Mackie's conceptual claim, which focuses on the concept of moral value, is framed in rationalist terms. He tells us that to say that happiness is a moral value is to say that happiness has the feature of being an end absolutely prescribed by reason. But insofar as they go in for conceptual analysis at all, the concept most rationalists seek to analyze is not the concept of moral value, but rather the concept of moral obligation. Do things look any better if we restate Mackie's argument in terms of the concept of moral obligation, rather than the concept of moral value? It might initially be thought that they do.

Rationalists typically tell us that the concept of a moral obligation is the concept of a certain sort of reason for action. Thus, for example, if we assume that act utilitarianism is the correct theory of what we ought to do – and from here on, in deference to the earlier decision to focus on Sidgwick's analysis of value, I will assume this for ease of exposition – then many rationalists claim it follows from this that there is a reason, perhaps conclusive perhaps non-conclusive (for there are stronger and weaker forms of rationalism), for each of us to maximize happiness and minimize suffering. One question to ask is whether rationalists mightn't just stop at this point. Or is there meant to be something metaphysically and epistemologically queer about the concept of a reason for action as such? Do reasons for action presuppose the existence of Moorean non-natural qualities? (Relatedly, we might ask whether there is meant to be something especially queer about *moral* obligation, or *moral* reasons, or whether the queerness is supposed to attach to the idea of there being anything at all that we ought to do, or any reasons for action at all.)

The answer is that Mackie's argument from queerness, if it works at all, establishes that the concept of a reason for action invoked by the rationalists is metaphysically and epistemologically queer as well. There are various ways to bring this out, but for present purposes the following observation should suffice. If we add to rationalism – understood as a claim about the link between moral obligations and reasons for action – the following “Williams-Korsgaard” thesis about the nature of reasons for action (Williams 1981; Korsgaard 1986):

WK: If an agent has a reason to ϕ then she would want herself to ϕ if she engaged in a suitable course of deliberation,

then it follows that claims about what we morally ought to do entail claims about what we would desire ourselves to do after a suitable course of deliberation. But since a suitable course of deliberation is simply one in which the deliberator is maximally informed and then forms his desires on the basis of that information in accordance with the requirements of rationality – this is something about which both Bernard Williams (a Humean) and Christine Korsgaard (a Kantian) agree – it further follows that claims about what we morally ought to do entail claims about what we would want ourselves to do after forming our desires in the light of full information and the requirements of rationality. (See also Smith 1994.) Mackie's

two challenges thus re-emerge, but this time about the possibility that our concept of a moral obligation is instantiated.

Why do Mackie's two challenges re-emerge? They re-emerge because, if he is right that there is something metaphysically and epistemologically queer about the idea that happiness is an end absolutely prescribed by reason, then there is something equally metaphysically and epistemologically queer about supposing that desiring to maximize happiness and minimize suffering is absolutely prescribed by reason. This, after all, is what we must be supposing if we think that we would each desire ourselves to maximize happiness and minimize suffering if we formed our desires in the light of full information and the requirements of rationality. If the former presupposes the existence of Moorean non-natural qualities, then so must the latter. Given WK, the argument from queerness is thus equally an objection to the rationalist's traditional analysis of the concept of moral obligation in terms of the concept of a reason for action. (Indeed, this suggests that the argument from queerness equally calls into question not just the existence of moral obligations, but there being anything we ought to do and the existence of any reasons for action.)

Nor, for the record, should this be in the least surprising. For a natural way to understand what it is to have a reason for action is in terms of the value produced by the action that we have reason to perform. As Davidson puts it, an action we have reason to perform must have certain “desirability characteristics” (Davidson 1963). The WK condition that must be satisfied by our moral obligations, if our moral obligations give rise to reasons for action, is thus much the same as the condition that must be satisfied by the states of affairs brought about by our actions if those states of affairs have moral value, given the Sidgwickian account of what it is for something to have moral value. In each case what's required is that there is some desire, or end, absolutely prescribed by reason. In the case of moral value, the desire in question concerns a state of affairs. In the case of a reason for action, the desire in question concerns the action of bringing that state of affairs about. There's no surprise here given that it follows from the fact that we have a reason to do something that our doing that thing will bring about something of value.

This detour is, however, helpful, as it enables us to focus more clearly on how we might respond to Mackie's two challenges. First of all, remembering once again our simplifying assumption about what we are morally obliged to do, we must explain either what the world would have to be like or what the requirements of rationality would have to be like for it to be the case that we would each desire ourselves to maximize happiness and minimize suffering if we formed our desires in the light of full information and the requirements of rationality. Second, we must show that it is plausible to suppose that the facts about the world, or the requirements of rationality, are like that. And third, on the assumption that we can know what's of moral value or what we have reason to do, we must make explicit how it is that we are able to acquire knowledge of the relevant facts about the world and requirements of rationality.

In what follows I will describe and assess what I take to be the four main approaches to responding to Mackie's two challenges. The discussion will of necessity be incomplete, as each approach can be developed in different ways, some of

which are more promising than others. My main aim, however, is not to say the last word about any of these approaches, but rather to put all four on the table for further discussion. My own view, to anticipate, is that only the fourth has any chance of showing what's wrong with Mackie's argument. Though I am I tempted by the fourth approach, I must confess to a sense of unease about it.

3 The Instrumental Approach

According to what I will call the "Instrumental Approach," only one requirement of rationality governing the formation of desires is needed to underwrite the truth of the claim that we would each desire ourselves to maximize happiness and minimize suffering if we formed our desires in the light of full information and the requirements of rationality. The requirement is some variation on the following, fairly uncontroversial, requirement of means-ends rationality (Note that hereafter "RR" is short for "Reason requires that"):

ME: RR (if a subject has an intrinsic desire that p and a belief that he can bring about p by bringing about q , then he has an instrumental desire that he brings about q).

This is because all of the work is done by a crucial empirical fact, one that we would each come to appreciate if we had full information, about the means by which we will get what we intrinsically desire, no matter what we intrinsically desire. (For a related argument see Gauthier 1986.) Before we get to that, however, let's focus for a moment on ME itself.

If the Instrumental Approach is to succeed, then we will have to explain what sort of fact ME is and we will also have to explain how it is possible for us to come by knowledge of this fact, in so far as we have knowledge of moral values and reasons for action. My own view is that ME is best understood in much the same way as we understand claims like "It ought to be the case that knives cut well." This "ought"-claim derives from the metaphysically mundane fact that *knife* is a functional kind. Since what a knife is is something whose function is to cut, it follows that knives can be ordered according to how well they serve that function: Knives that cut serve that function better than knives that don't; knives that cut more efficiently serve that function better than knives that cut less efficiently; and so on. As I see things, the claim that it ought to be the case that knives cut well is simply an efficient way of saying that this raft of evaluative claims is true. No Moorean non-natural qualities are thus required to underwrite the truth of this claim.

Similarly, it seems to me that ME derives from the metaphysically mundane fact that intrinsic desires and means-end beliefs are psychological states possessed by agents, where the psychology of an agent is also a functional kind. The psychology of an agent is something whose function is, *inter alia*, to produce action. Psychologies too can thus be ordered according to how well they serve their function: Those in which intrinsic desires combine with means-end beliefs in the way required to produce action – for this is what happens when they combine in such a way as to produce an instrumental desire (indeed, in my view, instrumental desires just are intrinsic desires that have suitably combined with means-end beliefs (Smith

2004)) – better serve the function of the psychology of an agent than do those in which intrinsic desires and means-end beliefs do not combine in this way; psychologies in which intrinsic desires and means-end beliefs reliably combine in this way serve that function better than those in which they combine in this way albeit unreliably; and so on. ME, which is just an "ought"-claim about the relationship between desires and means-end beliefs, is simply an efficient way of saying that this raft of evaluative claims is true.

Understood in this way, ME states a fact that is as metaphysically mundane as the claim that knives ought to cut well, and our knowledge of ME, much as our knowledge of the claim that knives ought to cut well, is mundane knowledge too. Just as the "ought"-claim about knives implied nothing about Moorean non-natural qualities, neither does ME. There is, however, one crucial difference between claims about the function of a knife and claims about the function of the psychology of an agent. Since knives are a human invention, the function of a knife can be traced in some loose way to human purposes. The psychology of an agent, by contrast, is not a human invention, so its function cannot be traced to human purposes. So what does fix the function of the psychology of an agent? Human psychologies are, of course, the causal product of a process of evolution, so it might be thought that the function of the psychology of an agent is fixed by the contingencies of that process. But the fact that a human psychology is the product of a process of evolution is not an essential feature of a human psychology *as a psychology*. ME purports to tell us something about the proper functioning of every possible psychology of an agent, including those that spring into existence willy-nilly, not just something about the contingencies of an evolved human psychology. Indeed, a human psychology might evolve in ways that make it flout ME. So, to repeat, what does fix the function of the psychology of an agent?

My own view is that we must think of the psychology of an agent as a privileged kind in reality – a Lewisian elite property, if you like (Lewis 1984) – and that we must suppose that knowledge of the function of the psychology of an agent is purely speculative knowledge of a thing of that kind. We gain knowledge of the function of the psychology of an agent *a priori* by reflecting on the nature of psychology and agency, much as we gain knowledge of causation, persistence, freedom, and the like, by reflecting on their natures. At this point, my suggestion is, when we reflect on the nature of the psychology of an agent we learn that that is something that has a function captured, *inter alia*, by ME. A crucial question that will need to be addressed, in attempting to respond to Mackie's argument, is whether we need to have a richer account of what it is for the psychology of an agent to function properly or whether this is all that needs to be said.

With this account of the metaphysics and epistemology of ME in place, the Instrumental Approach holds that what's been said is all that needs to be said. It holds that we need to appeal to just one further fact in order to explain why we would each desire ourselves to maximize happiness and minimize suffering – again, remember our simplifying assumption – if we formed our desires in the light of full information and the requirements of rationality. The further fact in question is a crucial empirical fact: namely, that acting so as to maximize happiness and minimize

suffering is an all-purpose means to the satisfaction of *whatever* desires agents happen to have. This is what full information would reveal to us, or so the Instrumental Approach tells us. Morality is an all-purpose means to our ends. In that case, no matter what people desire, so long as they have full information, and hence true beliefs about means, and so long as they go on to form their instrumental desires in accordance with the requirements of means-ends rationality in the light of these true beliefs, they will acquire an instrumental desire to maximize happiness and minimize suffering.

What should we make of this response to Mackie's argument?

The Instrumental Approach certainly succeeds in showing that there is a lacuna in Mackie's argument. Whereas Mackie says that the existence of desires that are absolutely prescribed by reason would require the existence of Moorean non-natural qualities, the Instrumental Approach shows that no such qualities are required. All that's required is ME and a crucial empirical fact. Unfortunately, however, the alleged empirical fact – the fact, given our simplifying assumption, that maximizing happiness and minimizing suffering is an all-purpose means to the satisfaction of whatever desires anyone happens to have – seems to be no empirical fact at all. And nor would it help if we were to eschew the simplifying assumption and make different assumptions about what we are obliged to do, and hence different assumptions about what the all-purpose means to our ends are. For whatever we are in fact obliged to do, it seems not to be an empirical fact that our doing that is an all-purpose means to the satisfaction of whatever desires we happen to have.

Some may balk at this. Couldn't God see to it that doing what we are obliged to do is an all-purpose means to the satisfaction of our desires, whatever desires we happen to have? Couldn't he institute a set of rewards for doing what we are obliged to do and punishments for our failing to do what we are obliged to do, where these rewards and punishments are in turn a matter of our getting whatever it is that we happen to want or be averse to? If so, and if God exists and sets up such a system of rewards and punishments, then the Instrumental Approach shows that he would thereby have seen to it that there exist features that are both objective and prescriptive. I said earlier that Mackie's argument purports to show that there could only exist objectively prescriptive features if there were Moorean non-natural features, something we can literally make no sense of. But, it might be thought, we now see that his argument shows no such thing. For the existence of moral value would be equally secured by the existence of God, on the assumption that God can indeed set up a system of rewards and punishments as described. However I am not persuaded that this last crucial assumption is plausible.

Suppose (for reductio) that God exists and sets up the required system of rewards and punishments in some possible world. So far, so good. Now let's ask a question about that possible world. Had there been someone whose intrinsic desires are not satisfied by his doing what's morally obligatory, would he have had moral obligations? This is a legitimate question to ask, because moral obligations, if they exist at all, are possessed not just by actual people, but also by those who would have existed had things been otherwise. And the answer is surely that he would still have

had moral obligations. Imagine, for example, someone whose every intrinsic desire is for some X where X does not come about by means of whatever it is that we are imagining it is morally obligatory for him to do. That person, by hypothesis, has moral obligations. But, given our analysis of moral obligation, he too would have to be rationally required to desire to do whatever it is that we are imagining he has a moral obligation to do. Yet if ME is the only requirement of rationality governing the formation of desires, we have imagined his having desires that make this condition impossible to satisfy. If this is right then the upshot is that not even God can help the Instrumental Approach explain the existence of moral obligations.

Let me sum up. If we take the Instrumental Approach, then, though we don't commit ourselves to the existence of Moorean non-natural qualities, we do land ourselves with the error theory nonetheless. We land ourselves with the error theory because it is incoherent to suppose that our doing what we are morally obliged to do is an all-purpose means to the satisfaction of our desires whatever we happen to desire: That's what the possibility of someone with desires like those just described shows. But the failure of the Instrumental Approach teaches us an important lesson. Imagine that everyone who is obliged to maximize happiness and minimize suffering would indeed desire themselves to so act if they formed their desires in the light of full information and the requirements of rationality. In that case there would have to be something irrational about someone whose every intrinsic desire is for some X where X doesn't come about by means of whatever it is that we are imagining we have an obligation to do. We thus have no alternative but to suppose that such intrinsic desires themselves are subject to rational requirements. Thus, the question to which we require an answer, given our simplifying assumption, is whether we mightn't be rationally required to intrinsically desire that we maximize happiness and minimize suffering. The remaining approaches all proceed by attempting to answer this question in the affirmative without presupposing the existence of Moorean non-natural qualities.

4 The Universalization Approach

If our intrinsic desires themselves are subject to rational requirements, then there must be rational requirements beyond ME. But what might such requirements of rationality be like?

One possibility is that, in order to be rational, our intrinsic desires must meet certain formal constraints, formal constraints that weed out all but the desire to maximize happiness and minimize suffering and those intrinsic desires whose satisfaction is consistent with the satisfaction of such a desire. The most obvious such constraint is some variation on the following requirement of universalization:

- U: RR (if a subject has an intrinsic desire that *p*, then either *p* itself is suitably universal or the satisfaction of the desire that *p* is consistent with the satisfaction of desires whose contents are themselves suitably universal).

The rationale for U is not hard to provide. (For a related argument see Kant 1786/1948.) If there are any norms of rationality governing desires at all, then, since no particulars have a privileged status in the rational order of things, it must be possible to state the norms in purely universal terms. There is an obvious analogy here with laws of nature. Because laws of nature assign no particulars a privileged role in the causal order, it follows that it must be possible to state them too in purely universal terms. According to U, this means that, to be rational, our intrinsic desires must have contents that are themselves suitably universal – they must mention no particulars – or, at any rate, their satisfaction must be consistent with the satisfaction of desires whose contents are themselves suitably universal. (From here on I will omit this qualification.)

According to the Universalization Approach, however, there is a further argument that takes us from U to the conclusion that the only desires concerning particulars that are rational are those either derived from or consistent with the desire to maximize happiness and minimize suffering. But what exactly the further argument is supposed to be is very much a moot point. The best known argument for something like this conclusion is that given by R.M. Hare (1981). (Having said that, however, it must immediately be added that Hare himself didn't accept U, which purports to state a requirement of rationality. According to Hare, though universalization is a condition of *morality*, it is not a condition of *rationality*. Hare therefore rejects the Sidgwickian analysis of value in terms of certain desires being absolutely prescribed by reason; this is where he parts company with Mackie. Let's, however, leave Hare's reasons for rejecting the Sidgwickian analysis of value, an analysis which Mackie accepts, to one side.) Hare did, however, argue that the only intrinsic desire that is suitably universal is a desire much like the intrinsic desire that there is as much happiness as possible. According to Hare, the only intrinsic desire that passes the universalization test is the desire that there is as much desire satisfaction as possible.

But now look at what happens if we put U together with Hare's views about the power of universalization arguments. We get the conclusion that, if we have any intrinsic desires at all, then, if we formed our intrinsic desires in the light of full information and the requirements of rationality, we would all desire that we maximize desire satisfaction. It might therefore be thought that the Universalization Approach provides us with a response to Mackie's two challenges all by itself. U is hardly metaphysically queer, after all. Much like ME, it is a principle that tells us what must be the case for the psychology of an agent to function properly. According to U, the psychology of an agent functions properly only if the desires that issue in action are themselves suitably universal. No Moorean non-natural qualities there. Nor is there any epistemological queerness either – not, at any rate, if Hare's arguments succeed in showing that the only intrinsic desire that passes the universalization test is the desire that there be as much desire satisfaction as possible. As I said, however, what the argument is supposed to be that establishes this conclusion is very much a moot point. Let me briefly explain why.

Hare thinks, plausibly enough, that U would require us, in whatever situation we happen to find ourselves, to find something that we want to obtain in every possible situation identical in universal respects to this situation. Let's suppose that we find

ourselves in a situation, which we will call "S," in which there are three people interacting in a certain way. In S: we desire that *p*, a second person desires that *q*, and a third person desires that *r*. In figuring out what we can want to obtain in every possible situation identical in universal respects to S, Hare suggests that we need to put together three distinct desires concerning three distinct ways things could be that are none the less identical in universal respects to S. The first is our desire that *p* and it concerns S itself. The second is the desire that we would acquire if we were fully to imagine ourselves in the possible situation identical in universal respects to S, but in which we ourselves occupy the position of the second person. According to Hare, this is the desire that, in the possible situation in which we occupy the second person's position, *q*. And the third is the desire that we would acquire if we were to fully imagine ourselves in the possible situation identical in universal respects to S, but in which we occupy the position of the third person. According to Hare, this is the desire that, in the possible situation in which we occupy that position, *r*. The upshot, according to Hare, is that we have three conflicting desires concerning three possible situations all of which are identical to S in purely universal respects. In forming a desire for one thing to obtain in all of these situations, he thinks that there is therefore only one rational course, and that is to desire whatever will maximally satisfy our three conflicting desires concerning these situations. This in turn, he thinks, is equivalent to our desiring to maximally satisfy the desires of all three parties involved in S.

There are many things to say about this argument, but for present purposes it will suffice to focus on just one. To repeat, Hare thinks that if we were fully to imagine ourselves occupying the position of (say) the second person, who desires that *q*, then we would acquire the desire that, in the possible situation in which we occupy that position, *q*. There are two ways to understand what he has in mind. One is that there is a necessary connection between belief and desire: When we form the belief that there is some possible situation in which we desire that *q* then that entails that we also form the desire that, in that possible situation, *q*. But since it is plainly false that there is such a necessary connection between belief and desire – it is at least possible for someone to believe that there is a possible situation in which she desires that *p* while being indifferent or averse to *p*'s being the case in that situation – the argument, so interpreted, isn't very plausible. The other possibility is that Hare is positing a normative connection between belief and desire. He might be thinking that reason requires us, when we believe that there is some possible situation in which we desire that *q*, to desire that, in that possible situation, *q*. We may of course have the belief but lack the desire, but only at the cost of flouting the normative requirement. The trouble with this interpretation of Hare's argument, however, is that it posits a rational principle independent of U. Whereas U tells us that there is a normative constraint on the form of our desires – our desires must have contents that are suitably universal – this rational principle tells us that the contents of our beliefs put normative constraints on the contents our desires. Understood in this way, Hare's argument isn't a version of the Universalization Approach at all. It is a version of a quite different approach, the one that I will consider next.

This brief consideration of the Universalization Approach hardly establishes that every variation on that approach is flawed. Perhaps a more plausible version can be found in Kant's own much more extensively discussed version of the Universalization Approach (Kant 1786/1948), though the ink spilled explaining why Kant's various arguments don't work strongly suggest otherwise (see, most recently, Parfit forthcoming). My own view, however, is that this is unlikely to be so. When you look closely at them, all versions of the Universalization Approach seem to share the crucial feature of Hare's just identified. Though it may be plausible to suppose that universalization is a condition of rationality, all versions of the Universalization Approach appeal to something beyond mere universalization in order to establish that some particular intrinsic desire is rationally required. Mackie's two challenges thus remain. In these terms, his challenge is to dispel the sense that there is something metaphysically and epistemologically queer about the further thing to which appeal is made in such arguments. Perhaps universalization itself doesn't presuppose the existence of Moorean non-natural qualities, but the further thing to which appeal is made does.

5 The Reasons Approach

In *What We Owe To Each Other*, Thomas Scanlon explicitly rejects the idea that goodness is a Moorean non-natural quality. He claims that what it is for something to be of value is for there to be a reason to want it, or to appreciate it, or to have some other attitudinal response towards it, where the different attitudinal responses are markers of different kinds of value (Scanlon 1998). This is his well-known "buck-passing" account of value. On the plausible assumption that the existence of a reason to want something entails that wanting that thing is absolutely prescribed by reason, his account entails Sidgwick's.

Derek Parfit concurs and elaborates on the nature of the reasons that we have for wanting things when the value in question is intrinsic value (Parfit forthcoming). Parfit says that what it is for something to be intrinsically good is for that thing to have intrinsic features that provide us with reasons to want intrinsically that those very features be realized. This suggests the following alternative account of the principles of rationality governing our intrinsic desires.

Remembering once again our simplifying assumption, the idea would have to be that the intrinsic nature of the states of affairs in which there is as much happiness as possible provides us with a reason to intrinsically desire that that state of affairs obtains, and hence a reason to desire that we bring that state of affairs about. If this claim about the reasons that exist for intrinsically desiring is correct, then that in turn suggests that our intrinsic desires are subject to the following rational principle:

BD: RR (if a subject believes that a state of affairs has the intrinsic nature of that state of affairs in which there is as much happiness as possible, then he intrinsically desires that that state of affairs obtains).

The idea, in other words, is that subjects are rationally required to form the beliefs and desires that there are reasons for forming when they believe that those reasons obtain; they are, in other words, rationally required to be sensitive to such reasons as they can appreciate. Let's call this the "Reasons Approach."

How plausible is it to suppose that there are rational principles of the kind that the Reasons Approach posits? In answering this question we must remember that the question isn't whether BD itself is plausible. The question is the more general one whether rational principles with BD's form are plausible. In other words, the question is whether it is plausible to suppose that, for some p and some q , there is a basic rational principle of the form:

RR (if a subject believes that p then she has an intrinsic desire that q).

Moreover, it is important to emphasize that the question concerns *basic* rational principles of this form because, on plausible assumptions, we can derive at least one principle with the same form as BD from a mere commitment to the existence of rational principles governing desires. This would not, however, be a vindication of the Reasons Approach. Let me explain why.

Assuming that desires, like beliefs, are indeed subject to rational requirements, it follows that, no matter what form the rational requirements on desires take – whether the requirements are like those posited on the Instrumental Approach, the Universalization Approach, or the Reasons Approach – those who have the capacity to reflect on the rational standing of their desires may, as a result of their reflection, form beliefs about what they would desire if their desires conformed to such rational requirements. But now imagine someone who does indeed form such beliefs. Let's suppose he forms the belief that he would intrinsically desire that q if his desires conformed to the rational requirements to which they are subject. It seems that we thereby imagine someone who is under rational pressure either to acquire the intrinsic desire that q or to give up his belief that that is indeed what he would desire if his desires conformed to rational requirements. In other words, we seem thereby committed to supposing that the following is a further requirement of rationality:

RR (if a subject believes that she would intrinsically desire that q if her desires conformed to all rational requirements then she has an intrinsic desire that q),

where this principle has exactly the same form as BD; it is that instance where p is the proposition that she would intrinsically desire that q if her desires conformed to all rational requirements. Given that we think our desires are subject to some rational requirements or other, it therefore seems that we're committed to their being subject to an *additional* rational requirement: a requirement that our desires match our beliefs about what our desires should be, given those rational requirements (Smith 2001).

Nor is it surprising that our desires should be subject to such an additional requirement of rationality, for what we imagine, when we imagine someone who has the capacity to form beliefs about the rational standing of his desires, and indeed his beliefs too for that matter, is someone who doesn't just have beliefs and desires that are subject to rational requirements, but someone who can bring about what

are, by his own lights, rational improvements in his beliefs and desires, taken as a whole. This capacity to reflectively manage one's beliefs and desires is a distinctive rational capacity, one that presupposes rational requirements that govern the self-management itself, rational requirements with the same form as BD. But, of course, these rational requirements are not themselves basic in the sense we're after. They piggy-back on our commitment to more basic rational requirements that govern our beliefs and desires in the first place, those about which we form beliefs when we engage in the process of self-management.

We now have a new way of putting the question we asked initially. How plausible is it to suppose that a principle of the kind to which we are committed on the Reasons Approach – BD – is both true and explanatorily basic in the sense of not piggy-backing on our commitment to more basic rational requirements?

It might be thought that our discussion of the Universalization Approach already suggests a positive answer to this question. After all, when we described Hare's version of the argument from U to the conclusion that the only desire that is suitably universal is the desire that there is as much preference satisfaction as possible, we saw that his argument made crucial appeal to the following principle:

RR (if someone believes that there is some possible situation in which she desires that q , then she desires that, in that possible situation, q).

This principle, which has the same form as BD, tells us in effect that a certain consideration – the nature of the possible situation in which we desire that q – provides us with a reason to form a certain desire, namely, the desire that, in that situation, q . The fact that Hare's argument made crucial appeal to this principle was, I suggested, symptomatic of the failure of the Universalization Approach. That's because, as we can now see, he was really offering a version of the Reasons Approach. To the extent that his argument has any appeal at all, its appeal rests entirely on the assumption that the principle just described expresses an explanatorily basic requirement of rationality.

In fact, however, as we will shortly see, not only could this principle *not* be explanatorily basic, it seems doubtful that any such principle could be explanatorily basic. To see why this is so we need to reflect for a moment on what would have to be the case for such a principle to be explanatorily basic, so consider a slightly different case. Why should we suppose that the following is an explanatorily basic rational principle governing our beliefs?

B: RR (if someone believes that p and believes that if p then q , then she believes that q).

The answer to this question may seem to run parallel to what we have already said about BD. We should believe that this principle is true and explanatorily basic, we might say, because the facts that p and that if p then q are reasons to believe that q , and rationality is a matter of sensitivity to reasons; that is, it is a matter of forming the beliefs that we have reason to believe when we believe that those reasons obtain. But in fact the answer in the case of B has an extra and important feature over and above anything that we can say about the principle to which we are committed on the Reasons Approach.

Suppose someone asks why the facts that p and that if p then q are reasons for believing that q . I take it that this is a legitimate question and that we can answer it by saying something about the way in which the facts that p and that if p then q bear upon the truth of q . This is because what these considerations are reasons for is the attitude of *believing*, where believing is in turn an attitude whose nature we need to explain in terms of its having the aim of truth. The facts that p and that if p then q are reasons for *believing* q because, inter alia, belief aims at the truth and there are truth-supporting relations between p , if p then q , and q . If this is right, however, then the worry with BD can be stated rather simply. The worry is that we cannot say anything in defense of BD remotely similar to what we just said in defense of B.

Suppose we ask why the fact that a state of affairs has the intrinsic nature of a state of affairs in which there is as much happiness as possible is a reason for *intrinsically desiring* that that state of affairs obtains. If the answer to this question were to run in parallel to the answer we just gave in the case of reasons for belief, then we would have to answer it by appealing to the aim of desire. But what is the aim of desire? The aim of desire obviously isn't truth. That is the aim of belief. Nor would it help to suggest that the aim of desire is something like satisfaction, given our beliefs. That plainly won't help us explain why the fact that a state of affairs has the intrinsic nature of a state of affairs in which there is as much happiness as possible is a reason for intrinsically desiring that that state of affairs obtains. This, at any rate, is what we learned from the failure of the Instrumental Approach. The only answer that seems likely to do the required work is that the aim of desire is the good. But the trouble with this answer is that the Reasons Approach is itself derived from the Scanlon/Parfit buck-passing account of the good, an account according to which the good is simply that which there is reason to desire. There therefore isn't a good independent account of what there is reason to desire that could play the role of explaining what makes the considerations that are reasons to desire reasons to desire. It would be viciously circular to explain why a consideration is a reason for desiring in terms of the fact that the good is the aim of desire and then to immediately go on and explain the good in terms of what there is reason to desire.

The upshot, it seems to me, is that if we adopt the buck-passing account of the good then we simply cannot explain why the fact that a state of affairs has the intrinsic nature of a state of affairs in which there is as much happiness as possible is a reason for intrinsically desiring that that state of affairs obtains in a way that parallels the explanation that we give of why the considerations that are reasons for beliefs are reasons for beliefs. But if this is right then it turns out that the basic relation out of which the Reasons Approach is constructed – the relation of a consideration's being a reason for desiring – is a relation whose nature is (so far, at any rate) utterly opaque to us. We literally have no idea what would make one consideration rather than another a basic reason for *desiring*.

This discussion of the Reasons Approach helps us better understand Mackie's original metaphysical challenge to the existence of value. As I said earlier, Mackie says that if happiness is of value then it follows that happiness has the feature of being an end absolutely prescribed by reason, "the same for all minds." We can now see that Mackie's argument really poses a dilemma. If happiness is an end absolutely

prescribed by reason then, at least if we take the Reasons Approach, it follows that there must be some consideration that provides us all with a reason to desire happiness. But in order to understand what it would be for some consideration to be a reason to desire happiness we would have to have some independent conception of the good as the aim of desire. We might suppose that this is why Mackie thinks that we are led inevitably to Moorean non-naturalism about the good with its attendant metaphysical and epistemological difficulties. This is one horn of the dilemma. There is, however, another horn on which we deny the existence of Moorean non-natural qualities and follow instead the buck-passers in conceiving of the good as that which there is reason to desire. On this horn, however, the problem is that we can give no explanation at all of what it is for certain considerations, as distinct from others, to be reasons for *desiring* at all. On this horn, the idea of there being reasons for desiring literally makes no sense.

The Reasons Approach must therefore be rejected. This is not to say that we must reject BD, the principle to which we are committed on the Reasons Approach. BD may well be true, for all that's been said. But if BD is true, then that will be because we have derived it in some way from something else that is itself explanatorily more basic. But what else is there to derive it from?

6 The Constitutivist Approach

To avoid the error theory, it seems that we must explain why certain intrinsic desires are rationally required without presupposing that there are explanatorily basic reasons for desiring, since such reasons do presuppose the existence of Moorean non-natural qualities. But how might we do this?

The only untried possibility I can imagine is that we might suppose that certain desires are constitutive of being fully rational. In other words, remembering once again our simplifying assumption about the substance of morality, we might suppose that the following is an explanatorily basic principle of rationality:

C: RR (people intrinsically desire that there is as much happiness as possible).

Let's call this the "Constitutivist Approach." (Parfit considers views of this kind when he discusses the "critical" versions of the present aim theory (Parfit 1984).) To anticipate, it seems to me that we are better placed to respond to Mackie's metaphysical and epistemological challenges to the existence of value if we take the Constitutivist Approach than if we take any of the others.

As with the other approaches, C states a condition on the proper functioning of the psychology of an agent. It says, in effect, that if it is to function properly, an agent's psychology must include the desire that there is as much happiness as possible. C is thus metaphysically innocent. It does not presuppose the existence of any Moorean non-natural qualities and it doesn't presuppose unexplained reason relations either. C thus differs crucially from BD, the principle to which we are committed on the Reasons Approach. BD assumes that we can explain why a fully rational person would desire that there is as much happiness as possible in terms

of reasons that exist for desiring. BD thus falls foul of the need to explain what it would be for a consideration to be a reason for desiring: This is what gives rise to the need to appeal to Moorean non-natural qualities and the attendant metaphysical and epistemological queerness. Since C makes no such assumption, it avoids this charge of metaphysical and epistemological queerness.

Even so, note that C would, if it were true, explain the truth of BD. After all, if it is constitutive of being rational that we desire that there is as much happiness as possible then it follows that we are rationally required, when we believe that a state of affairs has the intrinsic nature of a state of affairs in which there is as much happiness as possible, to intrinsically desire that that state of affairs obtains. We might even express this by saying that the intrinsic nature of happiness provides us with a reason to intrinsically desire happiness. What's crucial, however, is that BD, so understood, would not be explanatorily basic. What's explanatorily basic is rather C itself: the claim that we are rationally required to desire happiness. But how plausible is it to suppose that this is so?

In answering this question it is once again important to remember that the question isn't whether this particular rational principle, C, is itself plausible. The question is the more general one whether it is plausible to suppose that, for some *p*, there is a rational principle of the form:

RR (people intrinsically desire that *p*).

Moreover, it is also important to remember that this is a metaphysical question and that epistemological questions are therefore orthogonal. The principle to which we are committed on the Constitutivist Approach is a principle that captures what the psychology of an agent has to be like if it is to function properly. We thus mustn't suppose that the principle purports to state some sort of obvious analytic truth about rationality. C itself may be no such thing. But the mere fact that it is far from obvious that C is true is neither here nor there given that C purports to be a metaphysical thesis, rather than an analytic truth about rationality. To be sure, if C is true then it is something that we can discover simply by thinking about what the psychology of an agent is like if it is to function properly, but it may be difficult and non-obvious for all that.

So let's now face the question fairly and squarely. How plausible is it to suppose that there are rational principles like those to which we are committed on the Constitutivist Approach? Here is where matters get tricky. Those who urge the Constitutivist Approach upon us will insist that, to the extent that we are convinced by Mackie's conceptual claim, we just have to admit that each and every moral judgment we make commits us to a corresponding judgment that some desire or other is constitutive of being rational. The judgment that it is morally obligatory to keep some promise in certain specific circumstances, for example, commits us to the judgment that desiring to keep that promise in those specific circumstances is the product of some intrinsic desire that is constitutive of being rational plus facts about how keeping that promise leads to the satisfaction of that intrinsic desire; the judgment that it is morally obligatory to return a borrowed book in certain specific circumstances commits us to the judgment that the desire that we return the

borrowed book in those circumstances is the product of some intrinsic desire that is constitutive of being rational plus facts about how returning that book leads to the satisfaction of that intrinsic desire; and so we could go on. It thus seems that our moral beliefs commit us to the conclusion that there are rational principles of the kind to which we are committed on the Constitutivist Approach.

The question, however, is whether we can rest content with this commitment. Mackie might say that we cannot. The commitment to these principles is, he might say, grist for his mill. Since we know, *ex ante*, that no desires are constitutive of being rational, the upshot is thus that we have to do a *modus tollens* and give up our moral beliefs. But the other possibility, of course, is that we can indeed rest content with the commitment. On this alternative way of thinking, the response we just imagined Mackie giving begs the question. For even if we were disposed to think that no desires are constitutive of being rational *ex ante*, after being convinced that our moral beliefs commit us to the conclusion that there are such desires, and after seeing that the supposition that there are such desires is metaphysically innocent, we should simply revise that belief. Our moral beliefs commit us to the conclusion that there are rational principles of the kind to which we are committed on the Constitutivist Approach, so that's that.

I must confess that I find it difficult to say which of these two responses is correct. Should we think that one way of figuring out what the psychology of an agent is like when it functions properly is by engaging in ordinary moral reflection? If so, then we should conclude that ordinary moral reflection provides us with insight into which desires are constitutive of being rational. Or should we instead suppose that our *ex ante* beliefs about the nature of rationality are themselves true? If so then, if we are indeed disposed, *ex ante*, to deny that there are any desires that are constitutive of being rational, then we should conclude that our moral beliefs are all false. My somewhat tentative suggestion is that the first supposition is more plausible than the second. But I say this mindful of the fact that that I thereby merely express a hunch rather than the conclusion of a reasoned argument.

Once the suggestion that our moral views provide us with insight into the nature of rationality is on the table, an obvious way of figuring out which desires are constitutive of being rational presents itself. We figure out which desires are constitutive of being rational in exactly the same way as we figure out what the most fundamental moral principles are (compare Gilbert Harman on the autonomous approach to morality in Harman 1985). In other words, we should believe C, as opposed to some alternative claim about the desires constitutive of being rational, for much the same reason that we should believe that the principle of utility is the most fundamental moral principle, if indeed we should believe either of these things at all.

We figure out what the most fundamental moral principles are by engaging in a certain canonical method of reasoning: the process of reflective equilibrium (Rawls 1951; Daniels 1979). We try to get our considered judgments about what our moral obligations are in specific cases into equilibrium with our reflective judgments about the most general moral principles that govern those specific cases. Similarly, the suggestion goes, we can figure out which desires are constitutive of being rational by getting our considered judgments about what reason requires us to desire in

specific cases – where, remember, our commitments about this can simply be read off from our moral judgments about the specific cases, as the one entails the other – into equilibrium with our reflective judgments about what the most general intrinsic desires constitutive of being rational are that stand behind these more specific desires.

Suppose that, via the process of reflective equilibrium, we are led to the conclusion that the intrinsic value of happiness and disvalue of suffering best explain and justify our more specific judgments of moral obligation. In that case it seems that that same process of reasoning will inevitably lead us to conclude that intrinsically desiring that there is as much happiness as possible is constitutive of being rational. The upshot is thus that, on the Constitutivist Approach, our ability to vindicate the truth of a candidate rational principle such as C, in the sense of being rationally justified in believing it to be true, goes hand in hand with our ability to provide a similar vindication of the principle of utility itself. Moral theorising and theorising about the nature of rationality are one and the same. What can we say to those who disagree with us about what our moral obligations are if we take the Constitutivist Approach? Let's suppose that they have the beliefs and desires of someone with a firm commitment to commonsense morality, whereas we have the beliefs and desires of a committed utilitarian. One thing we can say is that, as we see things, they lack the desire that is constitutive of being rational, whereas we possess that desire. But it is worth adding that the account we have just given of how we come by knowledge of what our obligations are and which desires are constitutive of being rational shows that they may be liable to a different kind of charge as well. Those who do not believe the principle of utility may be such that, if they were to engage in the reflective equilibrium process, they would come to the conclusion that the principle of utility is the most fundamental moral principle, and in that case we can criticize them for being irrational in the further sense of having epistemically unjustified beliefs. Moreover, given that the belief they would have, if their beliefs were epistemically justified, commits them to the conclusion that they would desire to maximize happiness and minimize suffering if they had the desires constitutive of being rational, it follows that such agents are liable to a further charge of irrationality as well. For if they had epistemically justified beliefs, rationality would be on the side of their having a matching desire to maximize happiness and minimize suffering. In this way the Constitutivist Approach allows that there may be many different grounds on which we might criticize those with whom we have moral disagreements.

And what can we say to someone who disagrees with us about the nature of rationality? Imagine someone who agrees that moral beliefs commit those who have them to the conclusion that certain desires are constitutive of being rational – to this extent they follow the Constitutivist Approach – but who then goes on to insist that, since there are no desires that are constitutive of being rational, it follows that all moral beliefs are false. Perhaps he goes on to add "That's why I have never had any moral beliefs." What are we to say to him? As I have already indicated, the only thing to say to such a person is that we quite reasonably take our moral beliefs to provide us with insights into the nature of rationality, insights that he evidently lacks.

Unfortunately, this means that we will be unable to convince him that his views about the nature of rationality are mistaken. But it could hardly be a requirement on any philosophical view that you have to be able to convince the arbitrary person that that view is correct. It is surely enough that we are able to convince ourselves of the reasonableness of our own view.

Let me sum up. The Constitutivist Approach seems to me to offer the most promising way of responding to Mackie's metaphysical and epistemological arguments for the error theory. Let's grant that Mackie is right that something is of moral value just in case desiring that thing is, as Sidgwick says, "absolutely prescribed by reason." It turns out that all that this requires is that desiring that thing is constitutive of being fully rational. Controversial though this claim is, the important point is that there is nothing metaphysically queer about it. It presupposes neither Moorean non-naturalism nor an unexplained reason relation. To be sure, we may not be able to convince everyone of the truth of this claim. But who would have thought that we could?

Acknowledgments An earlier draft of this chapter was given at the Symposium on Moral Rationalism at the annual Australasian Association of Philosophy Conference held in Melbourne in 2008. I am grateful for helpful comments made by my co-symposiasts – Charles Pigden and Francois and Laura Schroeter – and by members of the audience, especially Daniel Cohen, Simon Keller, Norva Lo, and Geoffrey Sayre-McCord. I am also especially grateful for written comments I received from Richard Joyce and Simon Kirchin. Though they saved me from many errors, I fear that all too many remain. Finally, I would like to thank the anonymous referee who suggested that I fix my tortured prose.

References

- Blackburn, S. 1993. Errors and the phenomenology of value. In his *Essays in quasi-realism*, 149–165. New York: Oxford University Press.
- Daniels, N. 1979. Wide reflective equilibrium and theory acceptance in ethics. *Journal of Philosophy* 76: 256–282.
- Davidson, D. 1963. Actions, reasons, and causes. *Journal of Philosophy* 60: 685–700.
- Garner, R. 1994. *Beyond morality*. Philadelphia: Temple University Press.
- Gauthier, D. 1986. *Morals by agreement*. Oxford: Clarendon Press.
- Hare, R. M. 1981. *Moral thinking*. Oxford: Oxford University Press.
- Harman, G. 1985. Is There a single true morality? In *Morality, reason, and truth*, eds. D. Copp and D. Zimmerman, 77–99. Totowa: Rowman and Allanheld.
- Joyce, R. 2001. *The myth of morality*. Cambridge: Cambridge University Press.
- Kant, I. 1786. *Groundwork of the metaphysics of morals*, Trans. H. J. Paton. 1948. London: Hutchinson and Company.
- Korsgaard, C. 1986. Skepticism about practical reason. *Journal of Philosophy* 83: 5–25.
- Lewis, D. 1984. Putnam's paradox. *Australasian Journal of Philosophy* 62: 221–223.
- Lewis, D. 1989. Dispositional theories of value. *Proceedings of the Aristotelian Society* 63(suppl.): 113–137.
- Mackie, J. L. 1977. *Ethics: Inventing right and wrong*. London: Penguin.
- McDowell, J. 1985. Values and secondary qualities. In *Morality and objectivity*, ed. T. Honderich, 110–129. London: Routledge and Kegan Paul.

- Moore, G. E. 1903. *Principia ethica*. Cambridge: Cambridge University Press.
- Parfit, D. 1984. *Reasons and persons*. Oxford: Clarendon Press.
- Parfit, D. Forthcoming. *On what matters*.
- Rawls, J. 1951. Outline of a decision procedure for ethics. *Philosophical Review* 60: 177–197.
- Scanlon, T. M. 1998. *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Shafer-Landau, R. 2003. *Moral realism: A defence*. Oxford: Oxford University Press.
- Smith, M. 1994. *The moral problem*. Oxford: Blackwell.
- Smith, M. 2001. The incoherence argument: Reply to Shafer-Landau. *Analysis* 61: 254–266.
- Smith, M. 2004. Instrumental desires, instrumental rationality. *Proceedings of the Aristotelian Society* 78(suppl.): 93–109.
- Smith, M. 2006. Is that all there is? *The Journal of Ethics* 10: 75–106.
- Williams, B. 1981. Internal and external reasons. In his *Moral luck*, 101–113. Cambridge: Cambridge University Press.