

## CHAPTER 1

---

# META-ETHICS

---

MICHAEL SMITH

### 1. INTRODUCTION

---

When we judge actions to be morally right or wrong, or people to be morally good or bad, or outcomes to be just or unjust, we engage in the practice of moral appraisal. But this practice raises all sorts of puzzling questions. Certain of these questions are familiar substantive moral questions. To which principles do people implicitly or explicitly commit themselves when they engage in moral appraisal? Are the principles consequentialist or deontological in character, or is moral thought less general and more particularistic? Should we be monists about the good or pluralists? Are the goods all neutrally characterizable, or are some egocentric? Is justice a matter of equality, or is it rather a matter of giving priority to the worst off? Questions of these kinds address issues in *normative ethics*. (For further discussion of issues in normative ethics, see the next chapter, by Julia Driver.)

These moral questions are substantive in the sense that answers to them provide us with specific normative recommendations about how to act or what to prefer. Though sceptics about normative ethics might think that such recommendations amount to little more than sounding off, Rawls's detailed description of the method that we ordinarily use in justifying our answers to these substantive questions in normative ethics—the method of reflective equilibrium—did much to quarantine this kind of scepticism (Rawls 1951).<sup>1</sup> Very roughly, Rawls's idea is that we begin

<sup>1</sup> For further developments of the idea of reflective equilibrium, see Rawls (1971) and Daniels (1979). For opposing views, see Hare (1971) and Singer (1974).

with those specific moral judgements about which we are most confident and try to provide them with a compelling and comprehensive underlying justification. We then amend our initial attempt to provide a comprehensive justification in the light of our specific judgements when we are more confident about our specific judgements; we amend our specific judgements in the light of the attempted comprehensive justification when we are more confident about the comprehensive justification; and so on and so forth until, eventually, we achieve a reflective equilibrium: a point at which we deem no further mutual adjustments to be required. Rawls's suggestion is that when we have achieved such a reflective equilibrium, we are in a position to have confidence not just in the resulting comprehensive justification and the specific judgements that they entail, but also about the results of using the comprehensive justification to give answers on those substantive normative issues about which we were, antecedently, less confident. (For further discussion of the epistemology of moral judgement, see Chapter 3, by Karen Jones.)

But though this might reassure us, to some extent at least, that normative ethics isn't simply a domain in which we sound off—the normative judgements that survive the reflective equilibrium procedure meet certain epistemic standards, after all—there is still plainly an unstated anti-sceptical assumption lurking in the background. In terms of Rawls's method of reflective equilibrium, the assumption is that there are no considerations that undermine our confidence in all of our moral judgements at once. However, this assumption can be questioned. For at the other extreme from familiar substantive questions in normative ethics lie much more general and abstract questions about the nature of moral judgements themselves. These are questions that we can at least attempt to answer while remaining neutral about substantive questions in normative ethics. However, they are also questions the answers to which have the potential to undermine our commitment to morality itself. Let me give some examples.

The main question of this kind that has preoccupied moral philosophers in recent years is whether, when we make moral judgements, we express beliefs about the way the world is morally, or instead express some sort of non-belief state, a desire (say) that the world be a certain way in non-moral respects. This is the all-important question that divides *cognitivists* from *non-cognitivists*.<sup>2</sup> This question is all-important because the answer we give to it reveals radically different ontological commitments. While the former view presupposes that in making moral judgements we commit ourselves to the existence of a distinctive realm of moral facts, the latter suggests that we have no such commitment. On the view that moral

judgements express desires, although we go in for talk of 'moral facts' and 'moral beliefs', such talk is simply a loose manner of speaking. Strictly speaking there are no moral facts and no moral beliefs at all. There are merely desires that gain expression in syntactically complex sentences, where expression must be understood to be exactly the same relation as holds between emotions and exclamations such as 'Boo!' and 'Hurray!' On the view that moral judgements express beliefs, by contrast, talk of moral facts and moral beliefs is no mere manner of speaking. Absent a realm of moral facts, all such beliefs are false (Mackie 1977).

This all-important question suggests a range of related questions. For example, we can ask what the constitutive features of moral judgements are that tell in favour of one or another answer to the central question. Which features make them appear belief-like and which make them appear desire-like?<sup>3</sup> If moral judgements express desires, we can ask how we are to understand the syntactically complex semantics of the sentences we use when we make moral judgements. Why do such sentences look for all the world like fact-stating sentences, sentences that are typically used to express beliefs, and nothing like the exclamations that are typically used to express emotions?<sup>4</sup> And if moral judgements express beliefs, we can ask whether any such beliefs are true, and, if so, what makes them true. For example, if there are moral facts, then what differentiates them from facts of other kinds? Do moral facts require the existence of properties alien to science? If so, then do we have any reason to believe in the existence of such properties?<sup>5</sup>

It should be clear how these questions connect back to the earlier discussion of reflective equilibrium as a method of justification in normative ethics. For if moral facts require the existence of properties alien to science, and if for general metaphysical reasons we have no reason to believe that there are any such properties, then these considerations look as if they might well undermine our confidence in *all* of our moral judgements at once. And if that happens, then further questions suggest themselves. We can ask whether this would undermine the entire point of going in for moral appraisal, or whether we could happily go on as before by (say) simply pretending that our moral judgements are true. In other words, we can ask whether morality is nothing more than a useful fiction. If so, what is the use of the fiction?<sup>6</sup>

<sup>3</sup> This is the issue that divides internalists and externalists. Internalists include Hare (1952), Nagel (1970), Blackburn (1984), McDowell (1985), Korsgaard (1986), and Smith (1994). Externalists include Frankena (1958), Foot (1972), Railton (1986), and Brink (1986).

<sup>4</sup> This is the issue that divides quasi-realists—non-cognitivists who think that they can explain why moral language looks for all the world as if it expresses belief—and their opponents. Quasi-realists include Blackburn (1984, 1993) and Gibbard (1990, 2003). Opponents of quasi-realism include Hale (1986, 1993), who focuses on Blackburn, and Dreier (1996, 1999), who focuses on Gibbard.

<sup>5</sup> This is the issue that divides naturalists from their opponents. Naturalists include Railton (1986, 1993a,b), Boyd (1988), Jackson (1998), and Sturgeon (2003). Those who oppose naturalism include Moore (1903), Dancy (1986), Wiggins (1993a,b), and Shafer-Landau (2003).

<sup>6</sup> For contrasting views, see Nietzsche (1887), Mackie (1977), and Joyce (2001). See also Wallace (forthcoming).

<sup>2</sup> Cognitivists include Moore (1903), Frankena (1958), Nagel (1970), Foot (1972), Harman (1977), Mackie (1977), McDowell (1979, 1981, 1985), Wiggins (1987, 1990–1), Sturgeon (1985), Boyd (1988), Brink (1984, 1989), Railton (1986), Dreier (1990), Dancy (1993), Smith (1994), Jackson and Pettit (1995), Jackson (1998), and Shafer-Landau (2003). Non-cognitivists include Ayer (1936), Stevenson (1944), Hare (1952, 1963, 1981), Blackburn (1984, 1993, 1998), Gibbard (1990, 2003), and Timmons (1998).

These quite general questions address issues in *meta-ethics*. As the list indicates, in meta-ethics the focus is not on the substance of morality—not on specific recommendations about how to act or what to prefer or the principles that guide such recommendations—but is rather on a range of interrelated semantic, metaphysical, psychological, and epistemological questions about the practice of making moral judgements themselves. Moreover, to repeat, these are questions that we seem able to raise and answer without paying too much regard to the substance of morality. They are therefore questions on which we might hope to make progress even while remaining ignorant of that substance. Indeed, it is an important part of the methodology of meta-ethics that we do not unnecessarily prejudge any substantive issues in normative ethics.

In recent years the arguments given in favour of different answers to the meta-ethical questions outlined above have become highly articulated.<sup>7</sup> This has not, however, brought about much in the way of agreement. Even positions that were once regarded as having been decisively refuted are now finding new adherents.<sup>8</sup> Moreover, as arguments for positions in meta-ethics have become increasingly more sophisticated and subtle, there has been a noticeable shift in the attention of some meta-ethicists.<sup>9</sup> The idea seems to be that meta-ethics has erred in taking the standard picture of human psychology for granted. Future advances in meta-ethics will, some seem to think, come from a more explicit and nuanced account of the nature of the psychological states and capacities that make it possible for agents to deliberate and act morally: that we need a better story about the nature of belief and desire and the ways in which each of these is regulated not just by the other but also, perhaps, by perception and emotion. (For further discussion of issues in moral psychology, see Chapter 4, by R. Jay Wallace.)

As I hope this sketch of the current state of play in meta-ethics makes clear, contemporary meta-ethics has reached a kind of plateau. My own hunch is that what is needed is not so much a new theory, or a new argument in favour of an old theory, as a pause while we all take stock. We need to remind ourselves of some of the more general underlying questions that hover in the background of meta-ethical discussions without always being made explicit, questions whose answers provide us with the more general orientation from which we give the answers we do to the standard meta-ethical questions. To be sure, some of these questions are psychological in nature. To this extent I agree with those meta-ethicists who think we should focus more explicitly on issues concerning the nature of our psychological

states and capacities. But there are other questions as well, and it is vitally important that we do not lose sight of them.

The aim of the current chapter is to bring some of these more general underlying questions into the foreground and then to suggest some answers. My hope is that, if we explicitly address these more general underlying questions, we will eventually find ourselves able to return to the standard questions already committed to answers, answers on which there might well be more prospect of agreement. Failing that, we might find that we agree about so much already that our residual disagreements are far less troubling.

## 2. WHY META-ETHICISTS SHOULD FOCUS ON META-LEVEL QUESTIONS ABOUT REASONS AND RATIONALITY

An initial question that hovers in the background of discussions in meta-ethics is why we bother thinking about meta-ethical questions in the first place, as opposed to meta-level questions about other normative domains. Coming up with an answer to this question forces us to address a fundamental issue about the normative force of moral claims.

It is uncontroversial that there are multiple systems of norms. There are norms of rationality, norms of morality, legal norms, norms of etiquette, professional codes of conduct, norms that govern games, and so on. Just as we can ask meta-ethical questions, we can ask meta-level questions in each of these other domains. For example, we can ask meta-legal questions about the nature of legal judgments. Are judgements about what the law is expressions of our beliefs about a domain of legal facts, or are they rather expressions of some non-belief state, our desires (say) about the way the world is to be in non-legal respects? If they express beliefs, then what would the truth-makers of legal judgments have to be like and are there any such truth-makers? And we could ask the same sorts of questions about normative judgements in each of the other systems as well.

Viewed from this perspective, meta-ethics is just one area in which we can raise the host of interrelated semantic, metaphysical, psychological, and epistemological questions described earlier. So the question is why we concern ourselves with these questions in meta-ethics as opposed to the similar questions that arise in meta-rationality, or meta-law, or meta-etiquette. One natural answer is that ethics purports to be, in some yet to be specified sense, a more basic normative system than the others. For example, the norms in each of these other domains, or the most

<sup>7</sup> See especially the naturalistic version of cognitivism defended by Jackson (1998) and the most recent version of non-cognitivism defended by Gibbard (2003).

<sup>8</sup> I have in mind the resuscitation of an older form of non-naturalism in Shafer-Landau (2003).

<sup>9</sup> See e.g. the papers in Cullity and Gaut (1997), Korsgaard (1996), Little (1997), Blackburn (1998), Scanlon (1998, ch. 1), Dancy (2000), Gibbard (2003), Arpaly (2003), Railton (2003, pt. III), Smith (2004a, pt. I), and D'Arms and Jacobson (forthcoming).

important of them, might be thought to reduce, *inter alia*, to moral norms. If this were so, then it would be clear why we should begin by answering meta-ethical questions rather than the meta-level questions in the other domains, for the only way we could fully answer the other meta-level questions would in that case be by first answering those same questions in meta-ethics. Moreover, it does seem that, on at least some views about these other normative systems, moral norms do enjoy this more basic status.

For example, a central issue in jurisprudence, the area in which meta-legal questions are addressed in so far as they are addressed at all, is whether the legal norms of a jurisdiction are simply disguised moral norms. Advocates of natural law insist that they are (Finnis 1980; Dworkin 1986). Advocates of legal positivism insist that they are not (Hart 1961; Raz 1979). Thus, for example, Ronald Dworkin holds that a certain norm is a legal norm of a jurisdiction just in case it is entailed by the set of moral principles that best justifies the practices of that jurisdiction. Put in terms of judgements, Dworkin's view is that legal judgments are a subclass of moral judgements: when we say that such-and-such is a law we thereby make a moral judgement. If Dworkin is right, then it follows that, in order to say whether in making legal judgments we express beliefs or desires, we must first know whether in making moral judgements we express beliefs or desires. The two answers are tied together, with the meta-ethical question being more basic.

The trouble, however, is that if this is a good reason for supposing that we should answer meta-ethical questions prior to answering questions in meta-law, then a similar argument shows that we should answer meta-level questions about rationality before we answer meta-ethical questions. For a central issue in meta-ethics is similarly whether moral claims are just disguised claims about what there is reason to do, or perhaps disguised claims about what it is rational to do (for the time being I will not distinguish these two questions). The issue, in other words, is whether moral norms reduce to norms of reason or rationality. Advocates of rationalism, like Kant (1786), hold that they do, whereas advocates of anti-rationalism, such as Hume (1740), insist that they do not.<sup>10</sup> But if the rationalists are right, then in order to say whether in making moral judgements we express beliefs or desires we must first know whether in making judgements about what there is reason to do or judgements about what is rational to do, we express beliefs or desires. The two answers are once again tied, but this time the meta-level questions about reason and rationality are more basic than the meta-ethical questions. We answer the meta-ethical questions only by first answering the meta-level questions about norms of reason and rationality.

There is another reason for thinking that we should focus initially on meta-level questions about reason and rationality, one that parallels a reason that is often

given within jurisprudence for supposing that legal norms reduce to moral norms. As advocates of natural law point out, the law purports to be *authoritative* in a way that other systems of norms do not. One attraction of Dworkin's suggestion that a certain norm is a legal norm of a jurisdiction just in case it is entailed by the set of moral principles that best justifies the legal practices of that jurisdiction, for example, is that it seems to capture this authoritative element of law. The authoritative element of law thus seems to tell in favour of a reduction of legal norms to moral norms (Goldsworthy 1990).<sup>11</sup> Other institutionalized systems of norms, norms such as norms of etiquette, are not similarly authoritative. Their normative force seems to reside entirely in the brutally coercive power of those who are in a position to extract compliance: in the case of etiquette, the social elite whose interests are served by the enforcement of such norms (Foot 1972).

Whatever the merits of this line of argument within jurisprudence, the crucial point is that a strikingly similar case can be made for supposing that moral norms reduce to norms of reason or rationality. For if moral norms don't reduce to norms of reason or rationality, then we must ask in what sense moral norms could be authoritative. Are we supposed to think that people are morally obliged to act in ways that they have no reason to act, or ways it would be irrational for them to act? That seems absurd, on the face of it. How can we blame people for doing what they have every reason to do? We can, of course, make sense of the suggestion if we model moral norms on (say) norms of etiquette. But that merely underscores the difficulty. For if it is unattractive to suppose that legal norms are brutally coercive like norms of etiquette, then it is all that much more unattractive to suppose that moral norms are brutally coercive. The claim of moral norms to be authoritative is, after all, that much more secure (Korsgaard 1996).

The upshot is that those who are interested in meta-ethics should adopt the working hypothesis that moral norms reduce to norms of reason or rationality. They should focus their attention, initially at any rate, on meta-level questions about norms of reason and rationality. Nor should this be surprising. For though there are many systems of norms, the most basic and authoritative norms are plainly the norms of reason and rationality. As creatures with beliefs and desires we are, as such, subject to norms of reason and rationality. Belief and desire are defined in terms of their functional roles, where these functional roles mirror the rational liaisons that these states enter into. Norms of reason and rationality are thus, quite literally, inescapable (Lewis 1974; Davidson 1984). If moral norms too are supposed to be basic and authoritative, then the most straightforward way for this to be so would be for them to reduce to norms of reason or rationality.

If this is right, we should welcome the fact that some meta-ethicists have turned their attention to issues in moral psychology. For in looking for a better story about

<sup>10</sup> A modern defence of rationalism can be found in Korsgaard (1996) and a modern defence of anti-rationalism can be found in Blackburn (1998).

<sup>11</sup> For more on this issue, see Postema (1996), Holton (1998), and Smith (1999).

the nature of belief and desire and the ways in which each is regulated by the other, these theorists willy-nilly address questions about the nature and scope of the norms of reason and rationality. The important point to emphasize, however, is that these theorists must keep a vivid sense of the meta-level questions that thereby arise. For example, are the claims that these theorists make about the substantive norms of reason and rationality expressions of belief or desire? What features of such judgements count in favour of one or another answer to this question? What are the ontological commitments of the different answers? Are these ontological commitments credible?

### 3. DEONTIC CONCEPTS VERSUS EVALUATIVE CONCEPTS

How should we proceed? Within normative domains in general, and in the domain of reason and rationality in particular, common sense distinguishes between two broad families of concepts. On the one hand there are *evaluative* concepts and on the other there are *deontic* concepts. The crucial task of the theorist is to say how these two families of concepts are related (Ross 1930; Brandt 1979; Rawls 1971). Let me first say something about this distinction. I will then explain why focusing on this distinction fixes how we should proceed.

It is easy enough to give lists of evaluative and deontic concepts. The concepts of the good and the bad, the desirable and the undesirable, the better and the worse, and so on, are evaluative concepts, and the concepts of the obligatory, the forbidden, and the permissible are deontic. The distinction is, in other words, a generalization of the well-known distinction between the right and the good. But though it is easy enough to provide these lists, it is much more difficult to give a precise characterization of the principle by which these lists get generated. On all accounts, however, the principle has something to do with the possibility of holding people responsible. Roughly speaking, those normative claims that entail the possibility of holding some agent responsible are deontic, whereas normative claims that do not entail such a possibility are evaluative.

Thus, for example, when people do something that they shouldn't do, it follows more or less immediately that they are candidates for being held responsible. There may be exemptions or excuses, so we may not hold them responsible in fact, but they are at least candidates. It was up to them to act in the relevant way, so the question of their responsibility arises. When someone or something is good or bad, by contrast, this question does not immediately arise. This is most obvious when

we consider things that are good and bad that are not states of agents: sunsets and flowers, for example. But even when we consider examples of goods and bads that are states of agents, it is important to note that no issue about anyone's responsibility for those things arises *immediately*. For example, someone's experiencing pain may be a bad thing, but it doesn't follow from this that the person experiencing pain is a candidate for responsibility for his own pain because we don't yet have any reason to think that his experiencing pain was up to him. I will have more to say about how we might make this distinction more precise in what follows. For the time being, however, this rough characterization will suffice.

Concepts of both these sorts are in play in the domain of reason and rationality. When we talk of what agents rationally should or should not believe or desire or do, we seem to be in the realm of the deontic. Those who are subject to rational requirements are candidates for being held responsible for their successes and failures to live up to their obligations. It is up to people to believe and desire what they should believe and desire. Likewise, when we talk of what agents have reason to do, we once again seem to be in the realm of the deontic. It follows from the fact that someone has a reason to do something that they must account for their acting or failing to act on their reason: it is, in some sense, up to them to do what they have reason to do.

But evaluative concepts are plainly in play in the domain of reason and rationality as well. When we ask agents to defend their claim to have a reason to act in a certain way, what we require of them is, as Anscombe (1957) and Davidson (1963) emphasize, that they specify some 'desirability characteristic' possessed by their action (see also McDowell 1978). They must tell us something good or desirable that they will thereby bring about, and here we seem to be in a realm with the potential, at any rate, to be purely evaluative. For many of the things that agents would mention when asked to provide such a desirability characteristic—their own pleasure or desire satisfaction, for example—look as if they could well come about completely by happenstance. For example, if pleasure and desire satisfaction are features of an agent's actions that make them good or desirable, then it looks as if they are equally features of random events that just happen to bring them about that make those events good or desirable. They are still good or desirable, but in that case no one would be accountable for bringing them about.

Armed with this common-sense distinction between evaluative and deontic concepts, we face an initial set of meta-level questions about how these concepts are to be defined. For example, within the family of evaluative concepts we can ask what the relationship is between something's being good or bad, on the one hand, and its being better or worse than something else, on the other. The answer to this question is, I take it, relatively straightforward, at least to a first approximation. Goodness and badness come in degrees and are arranged on a scale ordered by the better and worse relations. More goodness is better than less, some goodness is better than none, no goodness and no badness is better than some badness, and less

badness is better than more. But here we note something crucially important. For the fact that we can identify a part of the scale to do with goodness, and a part to do with badness, and a point in between these two parts that is neither good nor bad, suggests that the concepts of goodness and badness are prior to the concepts of better and worse. There is more information on the scale than mere information about the ordering in terms of better and worse.

Likewise, within the deontic, we can ask what the relationship is between something's being obligatory, forbidden, and permissible. The key in this case is, I take it, relatively straightforward. We could begin with the concept of the obligatory. An act that is forbidden is an act that we are obliged not to perform, and an act that is permissible is an act that we are not obliged not to perform. Or we could begin with the concept of the forbidden. An act that is obligatory is an act that we are forbidden not to perform, and an act that is permissible is an act that we are not forbidden to perform. Or we could begin with the concept of the permissible. An act that is forbidden is one that is not permissible, and acts that are obligatory are acts that are uniquely permissible. In the case of deontic concepts, then, it seems that we can take any concept we like to be prior.

Definitional questions about the relationship *between* evaluative concepts and deontic concepts are, however, much more difficult to adjudicate. Are evaluative concepts definitionally prior to deontic concepts? Or are deontic concepts definitionally prior to evaluative concepts? Or are the concepts interdefinable, with neither enjoying definitional priority over the other? Or do we have two completely independent families of concepts, with neither being definable in terms of the other? It is not at all obvious how we should answer these meta-level questions. Such meta-level questions would, however, seem to be of the first importance. For meta-level questions about the kind of mental state that we express when we make normative claims must be asked about those claims that employ the definitionally most basic normative concepts.

In order to see that this is so, imagine for a moment that evaluative concepts are definitionally basic, and that deontic concepts are defined in terms of evaluative concepts together with some straightforwardly descriptive concepts. In that case, if we were to focus on the normative claims we make that are framed in terms of deontic concepts, what we would discover is that there is at least a straightforwardly factual element, an element expressive of belief, namely, that element specified in terms of straightforwardly descriptive concepts. But it would be wrong to conclude on this basis that normative claims in the domain of reason and rationality as such are expressive of belief. For the crucial question would still be left unanswered. Is the most basic element, the element framed in terms of evaluative concepts, also expressive of belief? Only confusion will result from asking meta-level questions about anything but the definitionally most basic normative concepts.

It is now clear how we should proceed. We should proceed by asking whether, within the normative domain of reason and rationality, evaluative concepts or

deontic concepts are definitionally prior. Since to my knowledge this question hasn't ever been explicitly addressed with regard to this particular domain—indeed, to my knowledge the question hasn't been explicitly formulated—we will take our lead from similar discussions in the moral domain.

## 4. MOORE'S DEFINITION OF THE DEONTIC

In *Principia Ethica* G. E. Moore (1903) argues that there is the following analytic connection between the fact that an act is our duty—an act that we ought to perform—and facts about the goodness and badness of that action's outcome.<sup>12</sup>

(x)(x ought to  $\phi$  in circumstances C iff  $\phi$ -ing is the unique action of those that x can perform in C that has the best outcome).

Moore's suggestion readily adapts to the normative domain of reasons and rationality. The idea would be that an agent has all-things-considered reason to  $\phi$  in circumstances C if and only if  $\phi$ -ing is the unique action of those she can perform in C that has the best outcome, where outcomes in turn are ranked in terms of their possession of various desirability characteristics.

Given Moore's definition of 'ought', note that we can readily define what it is for someone's acting in a certain way to be permissible (this is a matter of her acting in that way's being one among several acts she can perform that has the best outcome), and we can also readily define what it is for her acting in a certain way to be forbidden (this is a matter of the action's being one she can perform that has less than the best outcome). According to Moore, there are thus two crucial elements in the definition of deontic concepts. One element is evaluative: being obligatory, permissible, and forbidden are all a matter of an act's standing in a certain relation to the *best* outcome. The other element concerns our capacities: we are obliged, permitted, and forbidden to act in certain ways only if *we can* act in those ways.

In this way Moore's definition purports to make it transparent why evaluative concepts and deontic concepts differ in the way we described earlier. It suggests that the application of a deontic concept entails the possibility of holding someone responsible because responsibility follows in the wake of having the *option* or the *capacity* to bring about a good outcome. If an agent has the option of bringing about a good outcome, rather than a bad outcome, then we expect her to bring about the good outcome: she is a candidate for praise if she succeeds and for blame

<sup>12</sup> Though he still thought that the connection was a priori, under the influence of Russell (1910), Moore (1942) ultimately gave up on the suggestion that the connection is analytic.



if she fails. But the mere application of an evaluative concept entails nothing about the object of evaluation having been the outcome of an action that was one of an agent's options, so no issue of responsibility arises. The fact that someone experiences pain is a bad thing, but it is not as such responsibility-implicating because there is so far nothing for the person to give an account of. The fact that someone brings about his own experience of pain when he had the option of acting so as not to have this experience, by contrast, is responsibility-implicating. This is, I think, a great virtue of the Moorean definition.

Moreover, Moore's definition trades on an apparent truism about the relationship between the desirability characteristics possessed by an agent's options and that agent's being a candidate for praise and blame. For suppose an agent has the option of acting in one of two ways and that one of these possesses more in the way of desirable characteristics than the other. Suppose further that she takes the option of producing (say) a less desirable outcome. It would seem to follow analytically that the agent is a candidate for blame, rather than praise. For what could possibly be said in favour of producing an outcome that is worse when you have the option of producing one that is better? If, as Anscombe and Davidson suggest, all reasons for action have the backing of desirability characteristics, then the answer is quite literally that nothing can be said. For no reason could be given that doesn't bring a desirability characteristic in its wake.

But Moore's definition has some other much less attractive features as well. Most obviously, since his definition entails that the evaluative cannot be spelled out in deontic terms, it follows that either the evaluative has to be defined in terms that are neither evaluative nor deontic—that is, in terms that are not normative at all—or that the evaluative is simply indefinable. Moore himself argued on the basis of the Open Question Argument that it is implausible to suppose that the evaluative is definable in non-normative terms, so he drew the conclusion that the evaluative is indefinable. But few were prepared to follow him down this path (Darwall *et al.* 1992).

Moore's definition has another unattractive feature. His definition is inconsistent with the truth of certain very standard views about the substance of the norms of reason and rationality, views that characterize what we have reason to do in ego-centric terms. Consider the view that agents each have all-things-considered reason to act so as to satisfy their own current desires. On the methodological assumption mentioned earlier that in answering meta-level questions about reason and rationality we should try as much as possible not to rule out any views about the substance of these norms, we would have to hope that this standard view can at least be formulated in Moorean terms. For only then could its truth be a matter of substantive debate. Unfortunately, however, this standard view looks as if it can't be given a coherent formulation if we accept Moore's definition of the deontic in terms of the evaluative.

Suppose we ask what A and B have all-things-considered reason to do. According to Moore's definition they should each perform the unique act of those available to

them that is best. The substantive view that agents have all-things-considered reason to satisfy their own current desires is a view about the nature of the good-making features of acts: an agent's acts are good to the extent that they satisfy their own current desires. In Anscombe's and Davidson's terms, the satisfaction of one's own current desires is the *desirable characteristic* possessed by an agent's actions (Anscombe 1957; Davidson 1963). So far, then, it seems that the view should be straightforwardly formulable in Moorean terms. We need simply to plug the substantive claim about what is desirable into the Moorean definition. Unfortunately, however, matters are not so straightforward.

If we accept Moore's definition, then how do we limit the good-making features of an agent's actions to the satisfaction of *that agent's own current* desires? After all, if the extent to which A's act at an earlier time satisfies A's earlier-current desires is a good-making feature of A's act at the earlier time, and the extent to which A's act at a later time satisfies A's later-current desires is a good-making feature of A's act at the later time, it would seem to follow that there is at least some goodness to be found in the satisfaction of A's earlier desires and some goodness to be found in the satisfaction of his later desires. To the extent that A at the earlier time acts so as to satisfy his own current desires, then, and not the desires that he will have later, it would seem to follow that he ignores something that he has to admit is good that he could bring about by his acts, namely, the satisfaction of his own later-current desires. It would therefore seem to be inconsistent to hold that the good-making features of an agent's actions could be limited to the satisfaction of that agent's own *current* desires.

Similarly, it seems inconsistent to suppose that the good-making features of an agent's actions could be limited to the satisfaction of *that agent's own* current desires. After all, if the extent to which A's act satisfies A's current desires is a good-making feature of A's act, and the extent to which B's act satisfies B's current desires is a good-making feature of B's act, then both A and B are in a position to recognize that there is at least some goodness in the satisfaction of A's current desires and some goodness in the satisfaction of B's current desires. To the extent that A and B act so as to satisfy their own current desires, and not the current desires that each other has as well, it would therefore seem to follow that they too ignore some of the things that are good that they could bring about by their acts, namely, in A's case, the satisfaction of B's current desires, and in B's case, the satisfaction of A's current desires.

The upshot is that, if we accept the Moorean definition, those who think that agents have all-things-considered reason to act so as to satisfy their own current desires seem to be guilty of focusing on only a *part* of the good, not *the whole* of the good. And since such a restricted focus is ad hoc, their view seems to suffer from a kind of incoherence. They are committed, by their own lights, to the quite different view that agents have all-things-considered reason to satisfy desires period, without regard to whose desires they are or when they are possessed: a form of preference utilitarianism. Moore embraced the general form of this argument wholeheartedly. He held that once we properly understand the definitional relations between deontic

and evaluative concepts, a thesis established at the meta-level, we see that this has direct implications for which substantive normative views are coherent and which are not. Indeed, he himself argued against egoism and in favour of utilitarianism on precisely these grounds.

It must be said that this is a powerful line of argument if it works. However, most people who contemplate this argument are so convinced that the conclusion is false that they take the argument to be a *reductio* of the premiss (Broad 1942). In other words, they conclude that since Moore's definition of the deontic in terms of the evaluative really does entail that no reasons need to be characterized in egocentric terms, this means that the evaluative must not be definitionally prior to the deontic. But this reaction to Moore's argument is in fact totally misconceived. There is a flaw in Moore's argument all right, but the flaw leaves his definition of the deontic in terms of the evaluative firmly intact. What is called into question is not Moore's definition of the deontic in terms of the evaluative, but rather his assumption that goodness is metaphysically simple.

## 5. SIDGWICK'S DEFINITION OF THE EVALUATIVE

In holding that the evaluative is definitionally prior to the deontic, Moore thereby committed himself to the view that goodness is not itself further analysable in deontic terms. But of course Moore also held the much more radical view that goodness is metaphysically simple: goodness is not only not further analysable in deontic terms, but not further analysable in any terms whatsoever. In this respect it is worth comparing Moore's view with one of the theories he was reacting to at the time he was writing.

According to Henry Sidgwick (1907), the good is not metaphysically simple, but is rather a matter of what we ought to desire (Hurka 2003). More precisely, the Sidgwickian definition of 'good' can be formulated in the following terms.

$(x)(t)(p \text{ is good}_{x,t} \text{ iff } x \text{ at } t \text{ ought to desire that } p).$

Similarly, according to the Sidgwickian definition of 'bad', something is bad just in case we ought to be averse to it.

$(x)(t)(p \text{ is bad}_{x,t} \text{ iff } x \text{ at } t \text{ ought to be averse to } p).$

On the surface the Sidgwickian definitions seem to reverse the direction of definitional priority. But that isn't quite accurate. For there are many deontic

concepts, and while Moore was defining one of these in terms of 'good', the Sidgwickian definition of 'good' appeals to another quite different concept.

Moore's definition, at least as we have interpreted it, specifies in evaluative terms what it means to say that an agent has all-things-considered reason to act in a certain way: an agent's reasons for action are defined in terms of the values of the outcomes of the actions that the agent can perform. The Sidgwickian definition, by contrast, defines what it is for something to be of value in terms of a quite different concept: something's being of value is defined in terms of those desires that are possessed by one whose psychology is ideal, that is, a psychology that meets all rational requirements and other ideals of reason. But the concept of *an agent's having all-things-considered reason to act in a certain way* and the concept of *an agent's psychology's meeting all rational requirements and ideals of reason* are quite different, though related, concepts. They are different but related concepts because, in a Moorean spirit, we can define the former in terms of the latter.

$(x)(t)(x \text{ at } t \text{ has all-things-considered reason to } \phi \text{ in circumstances } C \text{ iff } \phi\text{-ing is the unique action of those } x \text{ can perform at } t \text{ that brings about what } x \text{ would desire most happens in } C \text{ if his psychology met all rational requirements and ideals of reason})$

or, more simply:

$(x)(t)(x \text{ at } t \text{ ought to } \phi \text{ in circumstances } C \text{ iff } \phi\text{-ing is the unique action of those } x \text{ can perform at } t \text{ that brings about the best}_{x,t} \text{ outcome in } C).$

Somewhat surprisingly, the Sidgwickian definition of 'good' and the Moorean definition of 'ought' are therefore consistent with each other.

The Sidgwickian definition of 'good' should sound familiar, for it bears a striking similarity to a whole range of theories of value that have been much discussed over the years: the fitting attitude theory (Ewing 1947, 1959), the ideal observer theory (Firth 1952), views that insist that values are like secondary qualities (McDowell 1985; Wiggins 1987), different versions of the dispositional theory (Lewis 1989; Smith 1994), and response-dependent theories (Johnston 1989). In common with all of these theories, the Sidgwickian definition of 'good' is, in Scanlon's (1998) terms, a *buck-passing* theory (see also the various theories discussed in Rabinowicz and Rønnow-Rasmussen 2004). It entails that there is no metaphysically independent property of goodness, but rather that something's being good is a matter of its being an object of the desires that the subject would have if that subject had an idealized psychology. The definitional buck is in this way passed from the concept of goodness to the concept of an idealized psychology: the latter is the definitionally basic concept.<sup>13</sup>

<sup>13</sup> Scanlon says that the buck is passed from the concept of the good to the concept of a reason: that the concept of a reason is the definitionally basic concept. His idea is presumably that an idealized



Moreover, the Sidgwickian definition just given is illuminating because it brings out a crucial feature of buck-passing theories. If goodness is a matter of what ought to be desired—or, to put it more carefully, if intrinsic goodness is a matter of what ought to be intrinsically desired—then it follows that our ordinary talk of this or that's being good is potentially misleading. Though such talk makes it look as if goodness could be, as Moore thought it was, metaphysically simple, when we talk of a thing's being good we must, at least implicitly, be talking about that thing's being good *relative to people and times*, namely, all of those who ought, at those times, to desire that thing. The subscripts on 'good' in the definition are markers of this fact. So if we follow Sidgwick, it turns out that, contrary to Moore, goodness cannot be a simple property. Goodness has *structure* (Smith 2003).

This is important because the Moorean argument given earlier for the conclusion that it cannot be the case that each of us has all-things-considered reason to act so as to satisfy our own current desires can now be seen to be driven entirely by the assumption that goodness is metaphysically simple, and hence not structured in this way. For suppose that goodness is structured in the way that the Sidgwickian definition suggests. There is then no conceptual barrier to the following's being true:

$(x)(t)(x \text{ at } t \text{ ought to desire the satisfaction of the desires } x \text{ has at } t).$

In other words, it could well be the case that people ought to desire the satisfaction of their own current desires. But if this is right then it would follow that:

$(x)(t)(\text{the satisfaction of the desires that } x \text{ has at } t \text{ is good}_{x,t}).$

And if this is so then the Moorean argument considered earlier collapses.

Consider the argument given for insisting that agents must consider more than *their own* current desires. From the fact that the satisfaction of the desires that A has at  $t$  is  $\text{good}_{A,t}$  and the satisfaction of the desires that B has at  $t$  is  $\text{good}_{B,t}$ , we cannot derive the conclusion that there are two good things where we understand goodness as being relative to the same person and time in each case. There is just one thing that is  $\text{good}_{A,t}$ , namely the satisfaction of the desires that A has at  $t$ , and one thing that is  $\text{good}_{B,t}$ , namely the satisfaction of the desires that B has at  $t$ . To suppose otherwise is to imagine that we can move from the premiss that each person ought to desire the satisfaction of their own current desires to the conclusion that each person ought to desire the satisfaction of everyone's current desires. But this

psychology is simply a psychology in which one's desires are suitably sensitive to reasons. But this assumes, unjustifiably, that the only way in which desires could be subject to rational requirements is if these rational requirements were, in turn, characterizable in terms of a sensitivity to reasons (Smith forthcoming). But that is simply false. Requirements of means–end rationality, for example, are not characterizable in terms of a sensitivity to reasons (Smith 2004b). The definitional buck may therefore be passed to the concept of a rational requirement without thereby being passed to the concept of a reason.

move is substantive. It is not licensed merely by the meaning of 'ought', but by a substantive claim about what we ought to desire.

The argument given for supposing that agents must consider more than the satisfaction of their own *current* desires is similarly shown to be defective. For agents can perfectly consistently hold that their own current desire satisfaction is  $\text{good}_{\text{them,now}}$  and that their later desire satisfaction is  $\text{good}_{\text{them,now}}$  without holding that their later desire satisfaction is  $\text{good}_{\text{them,now}}$ . To suppose otherwise is to imagine that we can move from the premiss that each person ought to desire the satisfaction of their own current desires to the conclusion that each person ought to desire the satisfaction of their own future desires. But this is also a substantive claim. It is not licensed merely by the meaning of 'ought', but by a substantive claim about what we ought to desire. However, the fact that the nature of the good things is in this way dependent on substantive claims about what ought to be desired is completely invisible if you think, as Moore does, that goodness is metaphysically simple. It is therefore no surprise that Moore found his own argument so convincing.

But though the Sidgwickian definition in this way enables us to diagnose the flaw in the Moorean argument given earlier for the conclusion that it cannot be the case that each of us has all-things-considered reason to act so as to satisfy our own current desires, it is important to note that it does not itself commit us to the opposite conclusion either. It is perfectly consistent with the Sidgwickian definition of 'good' that:

$(x)(y)(t)(t')(x \text{ at } t \text{ ought to desire the satisfaction of the desires } y \text{ has at } t').$

In other words, it is perfectly consistent with the Sidgwickian definition that desire satisfaction is of value quite independently of whose desire it is that is satisfied or when that desire is possessed. The Sidgwickian definition of 'good', unlike the Moorean definition, would thus seem so far to be in accord with the methodological constraint mentioned earlier, the constraint that in taking views at the meta-level we do not thereby rule out any substantive views at the normative level. If good-making features require egocentric characterization, then one attraction of the Sidgwickian definition of 'good' is that it allows such characterizations. But it does not require that good-making features be given an egocentric characterization. This is an additional attraction of the Sidgwickian definition.

The fact that the Sidgwickian definition of 'good' in terms of 'ought' blocks the Moorean argument for the conclusion that no all-things-considered reason needs to be characterized in egocentric terms should be reason enough to prefer it to Moore's suggestion that goodness is metaphysically simple. But the Sidgwickian definition has more to recommend it than just that. For when Anscombe (1957) and Davidson (1963) say that agents who claim to have reasons to act in a certain way must be prepared to specify some desirable feature brought about by their acting in that way, it is just very plausible to interpret this as the demand that people explain why the thing that they claim to have reason to do is something that they

ought to desire: as the demand, in other words, that they would desire that thing if their psychology met rational requirements and other ideals of reason. This is both very plausible in its own right—undermining the claim that someone has reason to act in a certain way seems to be one and the same task as showing that their desire to do it doesn't fit into a coherent psychological profile (Williams 1980; Smith 1995)—and is made even more plausible because so understanding the demand explains why goodness and badness stand in the logical relationships in which they do. Let me explain this last idea.

We noted earlier that goods and bads come in degrees and are at opposite ends of a scale ordered by the 'better than' relation with neither good nor bad as its mid-point. We noted that everyone must come up with some sort of explanation of why goods and bads stand in these logical relationships to each other. Those who accept a definition of 'good' along Sidgwickian lines have a plausible explanation. The explanation is that goodness and badness stand in these logical relationships because desires and aversions—where aversion can be thought of as desiring the absence of something—are states with different strengths and can therefore be put in an isomorphic ordering with indifference as its mid-point. In this way the structure of goodness and badness can be seen as a shadow cast by the structure of desire and aversion.

Suppose, for example, that a subject *S* at a time *t* desires very strongly that *p*, a little less strongly that *q*, that he is indifferent to *r* neither desiring it to obtain nor desiring it not to obtain, that he has a weak desire that *u* not obtain, and that he desires very strongly that *v* not obtain. Now suppose further that these are all desires that the subject ought to have: they are part of a psychology that meets all rational requirements and other ideals of reason. The Sidgwickian then has the materials with which to explain the logical relationships between the goods and the bads. For the pattern of the subject's desires, given that these are all desires that the subject ought to have, together with the Sidgwickian definition of good and bad, entails that while *p* and *q* are both goods<sub>*S,t*</sub>, *p* is even better<sub>*S,t*</sub> than *q*; that *q* is better<sub>*S,t*</sub> than *r*, which is neither good<sub>*S,t*</sub> nor bad<sub>*S,t*</sub>; that *r* is better<sub>*S,t*</sub> than both *u* and *v*, which are both bad<sub>*S,t*</sub>; and that *v* is even worse<sub>*S,t*</sub> than *u*. If facts about goodness and badness are simply facts about the relative strength of the desires that are part of a psychology that meets all rational requirements and other ideals of reason, then it comes as no surprise at all that goodness and badness can be ordered in terms of better and worse with neither good nor bad as a mid-point.<sup>14</sup>

Finally, the Sidgwickian definition of 'good' makes it plain why facts about goods and bads are not responsibility-implicating in the way in which facts about reasons for action are responsibility-implicating. Agents are, after all, capable of desiring

<sup>14</sup> Here I assume that we can characterize indifference dispositionally in a way that makes it quite different from the psychological state of someone who simply hasn't given the matter any thought. For further discussion of this difficult issue, see Dreier (forthcoming).

things that they cannot bring about. So if agents ought to have such desires, then it follows that the class of good things is much broader than the class of things that an agent has reason to do: indeed, the class of things that the agent has reason to do is a subclass of the things that are good, namely, that subclass of good things that the agent can bring about. This, in turn, explains why facts about goods and bads *as such* are not responsibility-implicating. They are not responsibility-implicating because they may or may not be things that agents can bring about. In this way we can see that it may not be up to anyone to make things better as such, or to prevent things from getting worse as such. What is rather up to agents is to act on such reasons as they have: to take the option of making things better, or to prevent things from getting worse, when they have the option.

To sum up, though the Sidgwickian definition of 'good' is given in deontic terms, the deontic terms to which it appeals are very different from the deontic terms that Moore is trying to define with his definition of 'ought'. The Moorean definition of 'ought' and the Sidgwickian definition of 'good' can therefore be happily combined. This in turn means that the Sidgwickian definition of 'good' can be used to help diagnose the flaw in the argument we considered earlier from the Moorean definition of 'ought' to the conclusion that all reasons for action are characterizable in non-egocentric terms. Finally, as we have seen, the Sidgwickian definition of 'good' is immensely plausible in its own right and is made even more plausible by its ability to explain why goodness and badness are structured in the way they are and why facts about goodness and badness differ from facts about reasons for action in the way that they do.

The upshot is that we have now found the definitionally most basic concepts on which we need to focus at the meta-level in our discussion of norms of reason and rationality. The most basic normative judgement is the claim that an agent's psychology meets all rational requirements and ideals of reason. It is therefore time to ask the all-important meta-level question within this domain.

## 6. COGNITIVISM OR NON-COGNITIVISM?

Is the judgement that an agent's psychology meets all rational requirements and ideals of reason the expression of a belief or a desire? In other words, should we be cognitivists or non-cognitivists about judgements such as these? Let me begin by offering what I take to be the completely flat-footed answer to this question. To anticipate, the flat-footed answer supports cognitivism and the objections to be considered will all be found wanting. This means that, at the end of the day, the flat-footed answer stands.

It is immensely difficult to provide a comprehensive list of all of the rational requirements to which a psychology is subject or an account of what all of the ideals of reason are. Though most would agree that these requirements include the requirements and ideals of coherence, unity, and informedness, as soon as any concrete proposal is made about what these requirements amount to, the proposals are hotly contested (Gibbard 1990; Blackburn 1998). Even so, the arguments for and against any concrete proposal are a priori arguments—the arguments do not themselves turn on empirical claims—and this means that there is some structure within a psychology—some specific non-normative way in which a psychology has to be when it exhibits coherence, unity, and informedness—from which it follows, a priori, that a psychology that is so structured is a psychology that meets all rational requirements and ideals of reason. In arguing for specific proposals about what coherence, unity, and informedness amount to, after all, we thereby argue in favour of the significance of certain sorts of non-normative features over others. But since the judgement that a psychology that is so structured in non-normative terms is plainly the expression of a belief, indeed a belief with a metaphysically innocent content, and since it follows a priori from some such judgement that the psychology is one that meets all rational requirements and ideals of reason, it follows that the claim that a psychology meets all rational requirements and ideals of reason is itself an expression of belief, and, indeed, a belief with a metaphysically innocent content (compare Jackson 2003).

As I understand it, this argument might be resisted for one of two kinds of reason. The first calls into question the premiss that it follows a priori from some specific claim about the way in which a psychology is structured that that psychology is one that meets all rational requirements and ideals of reason.<sup>15</sup> After all, the objection goes, many very, very smart people have spent a lot of time thinking about what these requirements and ideals are, and, notwithstanding their best efforts, they have not yet been able to come to an agreement. Not only do Kantians and Humeans disagree about the rationality of specific violations of the categorical imperative, even hard-nosed decision theorists disagree with each other about the rationality of choosing one box rather than two boxes when faced with Newcomb's Problem (Campbell and Sowden 1985).<sup>16</sup> In order to suppose that this is a

<sup>15</sup> This is a version of the familiar argument from disagreement: see Mackie (1977) and Gibbard (1990). For earlier replies, see Brink (1984) and Nagel (1986).

<sup>16</sup> The classic statement is found in Nozick (1969). There are two players: a predictor and a chooser. The chooser is presented with two boxes: an open box containing \$1,000, and a closed box that contains either \$1 million or \$0 (he doesn't know which). The chooser must decide whether she wants to be given the contents of both boxes, or just the contents of the closed box. The chooser knows that the day before, the predictor predicted how she will choose. If he predicted that the chooser will take only the closed box, then he put \$1 million in the closed box. If he predicted that the chooser will take both boxes, he left that box empty. Suppose that the predictor has been 100 per cent accurate in his many predictions in similar cases in the past and that the chooser knows this. Should the chooser take just the one closed box or take both boxes?

disagreement about a matter of fact that is a priori accessible, we must therefore suppose either that those who have been arguing about such matters for years have yet to canvass the crucial arguments that would bring agreement about, or that they aren't really as smart as they seem to be: that in not being convinced by some argument on offer that is, by hypothesis, convincing, some of them are being irrational. In other words, some of them are thereby shown to have psychologies that are themselves inappropriately structured.

But, the suggestion goes, neither of these hypotheses is credible. We seem to have canvassed all of the arguments and the people doing the canvassing really do seem to be very, very smart. So the only credible conclusion for us to draw is that the move from a specific claim about the way in which a psychology is structured to the conclusion that that psychology is one that meets all rational requirements and ideals of reason is not a priori, but is rather a matter of decision. We make the specific claim and then, if we have the relevant kind of desire that people have psychologies that are so structured, we express our desire in the form of a judgement to the effect that psychologies of that kind meet all rational requirements and ideals of reason. Those who lack such a desire are not disposed to make that judgement. But this reveals no failure in their ability to reason in an a priori manner. In this way we find that we are committed to a non-cognitivist account of the definitionally most basic concepts in the domain of norms of reason and rationality, not a cognitivist account.

However, the trouble with this response is that it too easily overgeneralizes. If this really were a good reason to be a non-cognitivist about judgements to the effect that certain psychologies meet all rational requirements and ideals of reason, it would be an equally good reason to be a non-cognitivist about large parts of philosophy itself: all of those parts on which smart philosophers disagree. I take that to be a *reductio*. More to the point, it simply isn't implausible to suppose either that we have yet to discover some crucial argument in favour of the claim that a psychology that is structured in a certain way is one that meets the requirements and ideals of reason, or that some of those whom we deem to be very, very smart philosophers are in fact incapable of appreciating certain sorts of arguments. Indeed, I would have thought that professional philosophers are all too vividly aware of the possibility that they are themselves incapable of appreciating the very arguments that they should find compelling.

The second reason that might be given for resisting the flat-footed argument just given is that, if the judgement that a psychology that exhibits a certain kind of structure thereby meets all requirements and ideals of reason were the expression of a belief, then this would make it mysterious how judgements of that kind are able to play the role that they play in our psychological economy.<sup>17</sup> After all, the

<sup>17</sup> This is a version of the familiar motivation argument: see Hume (1740: 456–8) and Blackburn (1984, ch. 6).

suggestion goes, those who make such judgements are disposed, in so far as they are rational, to make their psychologies exhibit that kind of structure. Yet how could this be so unless their judgements were the expression of a psychological state that is itself capable of playing this motivational role? Desires are the obvious kind of psychological state that are capable of playing this motivational role. So we should suppose that those judgements express desires, not beliefs. This provides us with a second independent reason for giving a non-cognitivist account of the definitionally most basic concepts in the domain of norms of reason and rationality.

But the trouble with this response is that it reveals a radically mistaken view of what it is to be a rational creature. No one should suppose that *desires* play the crucial role of enabling psychologies to evolve so as to meet requirements and ideals of reason. For imagine someone who desires a certain end and believes that acting in a particular way is a way of achieving that end. His having this desire and means–end belief is not enough to guarantee that he has a desire for the means, because he may be means–end irrational: his psychology may not have evolved so as to meet the requirements of instrumental rationality. His desire and means–end belief must be put together in the way required for desiring the means. But how is this putting together of desire and means–end belief accomplished? Let's suppose, for *reductio*, that what's needed is a desire to have a psychology that meets requirements and ideals of reason. How would having this desire help? For it too must work in the normal means–end way. The person we have imagined must therefore not only desire to have a psychology that meets requirements and ideals of reason, but he must also believe that, since he desires a certain end and believes that a particular means is a means to that end, so having a desire for that means would give him a psychology that meets requirements and ideals of reason, and then . . . and then what? The possession of this particular desire and means–end belief is not enough to guarantee that the person we are imagining desires the means to this particular end either. In other words, we still have no guarantee that his psychology evolves so as to meet the requirements of instrumental rationality. His desire to have a psychology that meets requirements and ideals of reason and his belief to the effect that desiring the means would give him a psychology that meets requirements and ideals of reason must themselves be put together in the way required for desiring the means. But how is the putting together of this particular desire and means–end belief accomplished? If we suppose that the person must have a further desire, then we are off on an infinite regress.

The proper conclusion to draw is therefore that we have misconceived the way in which psychologies evolve to meet requirements and ideals of reason. This is not underwritten by a desire to have a certain sort of psychology. Rather, it is underwritten by a distinct capacity to have a psychology that meets requirements and ideals of reason, a capacity that is of a piece with the kinds of inferential capacities to which Lewis Carroll drew our attention in his famous discussion of Achilles and the Tortoise (1895). Agents have and exercise such capacities in a way that requires

no mediation by beliefs about the manner of their own exercise. So even if we concede that those who judge that a psychology that exhibits a certain kind of structure thereby meets all requirements and ideals of reason are disposed, in so far as they are rational, to make their psychologies exhibit that kind of structure, this would not tell in favour of their judgement's being the expression of a desire. Their judgement may, for all that, be the expression of a desire, but it may also be the expression of a belief instead. The crucial point is simply that *whatever* psychological state their judgement expresses, its role in our psychological economy is secured by the agents' possession and exercise of the capacity to have a psychology that meets the requirements and ideals of reason. It is this capacity that they have in so far as they are rational, not a desire to have a certain sort of psychology.

The upshot is that we have been given no good reason to doubt the completely flat-footed reason given at the outset for supposing that the judgement that an agent's psychology meets all rational requirements and ideals of reason expresses a belief rather than a desire. Moreover, and just as importantly, the flat-footed reason we gave for this conclusion didn't require us to deny the obvious fact that it is hotly contested what exactly these requirements and ideals are. Nothing that was said entails that there are any such requirements and ideals, of course. For all that's been said we might even be massively deceived about the very coherence of the concept of requirements and ideals of reason. But, absent some concrete reason to think that these concepts are incoherent, it seems that we can in good conscience embrace cognitivism about such judgements and dismiss the idea that we are massively in error. Moreover, since we can use these concepts to give a Sidgwickian definition of the good, and then use our Sidgwickian concept of the good to give a Moorean definition of an all-things-considered reason to act in a certain way, it follows that we can in good conscience embrace cognitivism about these judgements and dismiss the idea that we are massively in error about them too.

## 7. CONCLUSION

The suggestion made at the outset was that meta-ethicists should adopt the working hypothesis that moral norms reduce to norms of reason and rationality. The payoff of doing so is, I hope, now apparent. For we have seen good reason to be cognitivists about normative judgements within the domain of norms of reason and rationality. If moral norms reduce to norms of reason and rationality, then it seems that we have equally good reason to be cognitivists about moral judgements as well.

Moreover, we have seen that judgements about what we have all-things-considered reason to do, understood in the Moorean terms in which we have understood them, have a familiar consequentialist structure. This in turn suggests that, when we reduce moral norms to norms of reason and rationality, that consequentialist structure will be preserved. Meta-ethicists should therefore adopt not just the working hypothesis that moral norms reduce to norms of reason and rationality, but that all moral theories are, at bottom, forms of consequentialism. Given our Sidgwickian understanding of goodness and badness, the kind of consequentialism on offer is of course consistent with some, or even all, goods and bads being ego-centric. To this extent the consequentialism on offer is quite different from the standard forms of consequentialism discussed in the normative ethics literature, for according to these theories goods and bads are all neutral. But the conclusion is still a striking one, one which promises all sorts of advantages within normative ethics (Dreier 1993; Jackson and Smith forthcoming).

Some dissenters will no doubt want to take issue with one or another of the claims on which we have built our meta-ethical defence of cognitivism and consequentialism. But the theoretical reasons for accepting these claims—the comprehensive, systematic, and unified picture of the twin normative domains of morality and reason to which they give rise, and the benefits that would come from a consequentialist regimentation of the various theories on offer in normative ethics—not to mention the arguments given in their favour, should at least give such dissenters pause for thought.

## REFERENCES

- Anscombe, G. E. M. (1957). *Intention*. Oxford: Blackwell.
- Arpaly, Nomy (2003). *Unprincipled Virtue*. New York: Oxford University Press.
- Ayer, A. J. (1936). *Language, Truth and Logic*. London: Gollancz.
- Blackburn, Simon (1984). *Spreading the Word*. Oxford: Oxford University Press.
- (1993). *Essays in Quasi-Realism*. Oxford: Oxford University Press.
- (1998). *Ruling Passions*. Oxford: Oxford University Press.
- Boyd, Richard (1988). 'How to Be a Moral Realist', in Geoffrey Sayre McCord (ed.), *Essays on Moral Realism*. Ithaca, NY: Cornell University Press.
- Brandt, Richard (1979). *A Theory of the Right and the Good*. Oxford: Clarendon Press.
- Brink, David O. (1984). 'Moral Realism and the Sceptical Arguments from Disagreement and Queerness', *Australasian Journal of Philosophy*, 62: 111–25.
- (1986). 'Externalist Moral Realism', *Southern Journal of Philosophy*, suppl. vol., 24: 23–42.
- (1989). *Moral Realism and the Foundations of Ethics*. New York: Cambridge University Press.
- Broad, C. D. (1942). 'Certain Features of Moore's Ethical Doctrines', in P. A. Schilpp (ed.), *The Philosophy of G. E. Moore*. Evanston, Ill.: Northwestern University Press.
- Campbell, Richmond, and Sowden, Lanning (eds.) (1985). *Paradoxes of Rationality and Cooperation*. Vancouver: University of British Columbia Press.
- Carroll, Lewis (1895). 'What the Tortoise Said to Achilles', *Mind*, 4: 278–80.
- Cullity, Garrett, and Gaut, Berys (eds.) (1997). *Ethics and Practical Reason*. Oxford: Oxford University Press.
- Dancy, Jonathan (1986). 'Two Conceptions of Moral Realism', *Proceedings of the Aristotelian Society*, suppl. vol., 60: 167–87.
- (1993). *Moral Reasons*. Oxford: Blackwell.
- (2000). *Practical Reality*. Oxford: Oxford University Press.
- Daniels, Norman (1979). 'Wide Reflective Equilibrium and Theory Acceptance in Ethics', *Journal of Philosophy*, 76: 256–82.
- D'Arms, Justin, and Jacobson, Daniel (forthcoming). 'Sensibility Theory and Projection', in David Copp (ed.), *Oxford Handbook of Ethical Theory*. Oxford: Oxford University Press.
- Darwall, Stephen, Gibbard, Allan, and Railton, Peter (1992). 'Toward *Fin de Siècle* Ethics: Some Trends', *Philosophical Review*, 101: 115–89.
- Davidson, Donald (1963). 'Actions, Reasons and Causes', repr. in Davidson, *Essays on Actions and Events*. Oxford: Oxford University Press, 1980.
- (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
- Dreier, James (1990). 'Internalism and Speaker Relativism', *Ethics*, 101: 6–26.
- (1993). 'Structures of Normative Theories', *The Monist*, 76: 22–40.
- (1996). 'Expressivist Embeddings and Minimalist Truth', *Philosophical Studies*, 83: 29–51.
- (1999). 'Transforming Expressivism', *Noûs*, 33: 558–72.
- (forthcoming). 'Negation for Expressivists', in Russ Shafer-Landau (ed.), *Oxford Studies in Metaethics*, i. Oxford: Oxford University Press.
- Dworkin, Ronald (1986). *Law's Empire*. Cambridge, Mass.: Harvard University Press.
- Ewing, A. C. (1947). *The Definition of Good*. London: Macmillan.
- (1959). *Second Thoughts in Moral Philosophy*. London: Routledge & Kegan Paul.
- Finnis, John (1980). *Natural Law and Natural Rights*. Oxford: Clarendon Press.
- Firth, Roderick (1952). 'Ethical Absolutism and the Ideal Observer', *Philosophy and Phenomenological Research*, 12: 317–45.
- Foot, Philippa (1972). 'Morality as a System of Hypothetical Imperatives', repr. in Foot, *Virtues and Vices*. Berkeley: University of California Press, 1978.
- Frankena, William (1958). 'Obligation and Motivation in Recent Moral Philosophy', in A. I. Melden (ed.), *Essays on Moral Philosophy*. Seattle: University of Washington Press.
- Gibbard, Allan (1990). *Wise Choices, Apt Feelings*. Oxford: Clarendon Press.
- (2003). *Thinking How to Live*. Cambridge, Mass.: Harvard University Press.
- Goldsworthy, Jeff (1990). 'The Self-Destruction of Legal Positivism', *Oxford Journal of Legal Studies*, 10: 449–86.
- Hale, Bob (1986). 'The Compleat Projectivist', *Philosophical Quarterly*, 36: 65–84.
- (1993). 'Can There Be a Logic of Attitudes?', in John Haldane and Crispin Wright (eds.), *Reality, Representation, and Projection*. Oxford: Oxford University Press.
- Hare, R. M. (1952). *The Language of Morals*. Oxford: Oxford University Press.
- (1963). *Freedom and Reason*. Oxford: Oxford University Press.
- (1971). 'The Argument from Received Opinion', in Hare, *Essays on Philosophical Method*. London: Macmillan.

- Hare, R. M. (1981). *Moral Thinking*. Oxford: Oxford University Press.
- Harman, Gilbert (1977). *The Nature of Morality*. New York: Oxford University Press.
- Hart, H. L. A. (1961). *The Concept of Law*. Oxford: Clarendon Press.
- Holton, Richard (1998). 'Positivism and the Internal Point of View', *Law and Philosophy*, 17: 597–625.
- Hume, David (1740). *A Treatise of Human Nature*, ed. L. A. Selby-Bigge. Oxford: Clarendon Press, 1968.
- Hurka, Thomas (2003). 'Moore in the Middle', *Ethics*, 113: 599–628.
- Jackson, Frank (1998). *From Metaphysics to Ethics*. Oxford: Oxford University Press.
- (2003). 'Cognitivism, A Priori Deduction, and Moore', *Ethics*, 113: 557–75.
- and Pettit, Philip (1995). 'Moral Functionalism and Moral Motivation', *Philosophical Quarterly*, 45: 20–40.
- and Smith, Michael (forthcoming). 'Absolutist Moral Theories and Uncertainty'.
- Johnston, Mark (1989). 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society*, suppl. vol., 63: 139–74.
- Joyce, Richard (2001). *The Myth of Morality*. Cambridge: Cambridge University Press.
- Kant, Immanuel (1786). *Groundwork of the Metaphysics of Morals*. London: Hutchinson.
- Korsgaard, Christine (1986). 'Scepticism About Practical Reason', *Journal of Philosophy*, 83: 5–25.
- (1996). *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Lewis, David (1974). 'Radical Interpretation', *Synthese*, 27: 331–44.
- (1989). 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society*, suppl. vol., 63: 113–37.
- Little, Margaret (1997). 'Virtue as Knowledge: Objections from the Philosophy of Mind', *Noûs*, 31: 59–79.
- McDowell, John (1978). 'Are Moral Requirements Hypothetical Imperatives?', *Proceedings of the Aristotelian Society*, suppl. vol., 52: 13–29.
- (1979). 'Virtue and Reason', *The Monist*, 62: 331–50.
- (1981). 'Non-Cognitivism and Rule-Following', in Steven Holtzman and Christopher Leich (eds.), *Wittgenstein: To Follow a Rule*. London: Routledge.
- (1985). 'Values and Secondary Qualities', in Ted Honderich (ed.), *Morality and Objectivity*. London: Routledge & Kegan Paul.
- Mackie, J. L. (1977). *Ethics: Inventing Right and Wrong*. Harmondsworth: Penguin.
- Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.
- (1942). 'A Reply to my Critics', in P. A. Schilpp (ed.), *The Philosophy of G. E. Moore*. Evanston, Ill.: Northwestern University.
- Nagel, Thomas (1970). *The Possibility of Altruism*. Princeton: Princeton University Press.
- (1986). *The View from Nowhere*. Oxford: Oxford University Press.
- Nietzsche, Friedrich (1887). *On the Genealogy of Morals/Ecce Homo*, ed. Walter Kaufmann, trans. Walter Kaufmann and R. J. Hollingdale. New York: Vintage Books, 1989.
- Nozick, Robert (1969). 'Newcomb's Problem and Two Principles of Choice', in Nicholas Rescher (ed.), *Essays in Honour of Carl G. Hempel*. Dordrecht: Reidel.
- Postema, Gerald J. (1996). 'Law's Autonomy and Public Practical Reason', in Robert P. George (ed.), *The Autonomy of Law: Essays on Legal Positivism*. Oxford: Clarendon Press.
- Rabinowicz, Woldek, and Rønnow-Rasmussen, Toni (2004). 'The Demon Strikes Back: On Fitting Pro-Attitudes and Value', *Ethics*, 114: 391–423.
- Railton, Peter (1986). 'Moral Realism', *Philosophical Review*, 95: 163–207.
- (1993a). 'What the Noncognitivist Helps Us to See the Naturalist Must Help Us to Explain', in John Haldane and Crispin Wright (eds.), *Reality, Representation, and Projection*. Oxford: Oxford University Press.
- (1993b). 'Reply to David Wiggins', in John Haldane and Crispin Wright (eds.), *Reality, Representation, and Projection*. Oxford: Oxford University Press.
- (1997). 'On the Hypothetical and Non-Hypothetical in Reasoning About Action', in Garrett Cullity and Berys Gaut (eds.), *Ethics and Practical Reason*. Oxford: Oxford University Press.
- (2003). *Facts, Values, and Norms*. New York: Cambridge University Press.
- Rawls, John (1951). 'Outline of a Decision Procedure for Ethics', *Philosophical Review*, 60: 177–97.
- (1971). *A Theory of Justice*. Cambridge, Mass.: Harvard University Press.
- Raz, Joseph (1979). *The Authority of Law*. Oxford: Clarendon Press.
- Ross, W. D. (1930). *The Right and the Good*. Oxford: Oxford University Press.
- Russell, B. (1910). 'The Elements in Ethics', repr. in Russell, *Philosophical Essays*. London: Routledge, 1944.
- Scanlon, T. M. (1998). *What we Owe to Each Other*. Cambridge, Mass.: Harvard University Press.
- Shafer-Landau, Russ (2003). *Moral Realism: A Defence*. Oxford: Oxford University Press.
- Sidgwick, Henry (1907). *The Methods of Ethics*. Indianapolis: Hackett, 1981.
- Singer, Peter (1974). 'Sidgwick and Reflective Equilibrium', *The Monist*, 58: 490–517.
- Smith, Michael (1994). *The Moral Problem*. Oxford: Blackwell.
- (1995). 'Internal Reasons', *Philosophy and Phenomenological Research*, 55: 109–31.
- (1999). 'Morality and Law', in Christopher B. Gray (ed.), *The Philosophy of Law: An Encyclopedia*. New York: Garland.
- (2003). 'Neutral and Relative Value After Moore', *Ethics*, 113: 576–98.
- (2004a). *Ethics and the A Priori: Selected Essays in Moral Psychology and Meta-Ethics*. New York: Cambridge University Press.
- (2004b). 'Instrumental Desires, Instrumental Rationality', *Proceedings of the Aristotelian Society*, suppl. vol., 78: 93–109.
- (forthcoming). 'Is there a Nexus Between Reasons and Rationality?', in Sergio Tenenbaum (ed.), *Poznan Studies in the Philosophy of Science and Humanities: New Trends in Moral Psychology*. Amsterdam: Rodopi.
- Stevenson, Charles (1944). *Ethics and Language*. New Haven: Yale University Press.
- Sturgeon, Nicholas (1985). 'Moral Explanations', in David Copp and David Zimmerman (eds.), *Morality, Reason and Truth*. Totowa, NJ: Rowman & Allanheld.
- (2003). 'Moore on Ethical Naturalism', *Ethics*, 113: 528–56.
- Timmons, Mark (1998). *Morality without Foundations*. New York: Oxford University Press.
- Wallace, R. Jay (forthcoming). 'Ressentiment, Value, and Self-Vindication: Making Sense of Nietzsche's Slave Revolt', in Brian Leiter and Neil Sinhababu (eds.), *Nietzsche and Morality*. Oxford: Oxford University Press.
- Wiggins, David (1987). 'A Sensible Subjectivism', in Wiggins, *Needs, Values, Truth*. Oxford: Blackwell.
- (1990–1). 'Moral Cognitivism, Moral Relativism, and Motivating Moral Belief', *Proceedings of the Aristotelian Society*, 91: 61–86.



- Wiggins, David (1993a). 'Cognitivism, Naturalism, and Normativity: A Reply to Peter Railton', in John Haldane and Crispin Wright (eds.), *Reality, Representation, and Projection*. Oxford: Oxford University Press.
- (1993b). 'A Neglected Position?', in John Haldane and Crispin Wright (eds.), *Reality, Representation, and Projection*. Oxford: Oxford University Press.
- Williams, Bernard (1980). 'Internal and External Reasons', repr. in Williams, *Moral Luck*. Cambridge: Cambridge University Press, 1981.