7

DESIRES, VALUES, REASONS, AND THE DUALISM OF PRACTICAL REASON

Michael Smith

Abstract

In *On What Matters* Derek Parfit argues that facts about reasons for action are grounded in facts about values and against the view that they are grounded in facts about the desires that subjects would have after fully informed and rational deliberation. I describe and evaluate Parfit's arguments for this value-based conception of reasons for action and find them wanting. I also assess his response to Sidgwick's suggestion that there is a Dualism of Practical Reason. Parfit seems not to notice that his preferred value-based conception of reasons for action augurs strongly in favour of a view like Sidgwick's.[1]

Derek Parfit's main task in the early chapters of *On What Matters* is to state and defend a version of a 'value-based' theory of reasons for action against various versions of what he calls a 'desire-based' theory.[2] As the names suggest, value based theories hold that an agent's reasons for action are a function of the values that can be realized by his actions. Desire-based theories, by contrast, hold that they are a function of the desires, perhaps idealized, that his actions will satisfy.

Value-based and desire-based theories are rival theories about what it is for something to be a reason for action. Another sort of question we can ask about reasons is which particular acts we have

[1] An earlier version of this chapter was read at a conference on Derek Parfit's book manuscript held at the University of Reading in 2006. My thanks go to all those who participated, especially Derek Parfit.

[2] All quotes in the text from *On What Matters* are taken from the 7 November 2007 version of the manuscript which was at that point called *Climbing the Mountain*. The 16 April 2008 version of the manuscript contains some important changes that I discuss below in footnote 15. There are also some less significant changes. For example, the distinction between desire-based and value-based theories of reasons for action becomes the distinction between 'Subjectivism about Reasons' and 'Objectivism about Reasons': same distinction, different names. Since it was agreed that the chapters in this book volume would refer to the 7 November 2007 version of the manuscript, I will stick to the earlier more descriptive names in what follows.

reason to perform in various circumstances. Theorists who hold quite different views about what makes something a reason for action might yet agree in their answers to this more practical question. (Compare: theorists who accept quite different meta-ethical theories – intuitionism and definitional naturalism (say) – might yet both agree that utilitarianism correctly characterizes what we morally ought to do.) Though Parfit postpones a full discussion of practical questions until later in the book, he does consider two practical claims in the early chapters. One is a claim that he takes to be a datum, namely, that we all have reasons to want to avoid future agony and thus to try to avoid it if we can. The other is a claim with which he partially agrees and partially disagrees, namely, Sidgwick's suggestion that, since we have both egoistic and impartial reasons for action, where these reasons are incommensurable, it follows that there is a dualism of practical reason.

Though there is much to agree with in the early chapters of *On What Matters*, I do have some misgivings. For one thing, the distinction Parfit makes between value-based and desire-based theories of reasons for action seems to me not to capture the crucial differences between rival theories. To capture these differences we need to make rather different distinctions. As we will see, the full import of this taxonomic point becomes clear when we consider Parfit's own preferred explanation of his datum. For another, Parfit's response to Sidgwick's arguments for the dualism of practical reason is difficult to understand. Indeed, given what Parfit tells us about the nature of reasons for action themselves, it seems to me that, notwithstanding his response to Sidgwick, he might well be committed to the incommensurability of reasons for action in an even more radical way than Sidgwick.

The chapter is in five sections. In Section 1 I describe and evaluate Parfit's account of desire-based theories of reasons for action. In Section 2 I explain his views about the nature of value. In Section 3 I describe and evaluate his account of value-based theories of reasons for action in the light of his views about the nature of value. In Section 4 I consider and respond to the arguments he gives for preferring value-based theories of reasons for action to desire-based theories. As we will see, his arguments for preferring value-based theories turn on his preferred explanation of the alleged datum that we all have a reason to want to avoid future agony. And then, finally, in Section 5 I consider Parfit's discussion of Sidgwick on the dualism of practical reason.

1. Desire-Based Theories of Reasons for Action

Parfit tells us that:

> According to one group of theories, all . . . reasons [for action] are provided by facts about what might fulfil or achieve our present telic desires or aims. Some of these theories appeal to our actual present desires or aims. Others appeal to the desires or aims that we would now have, or would want ourselves to have, if we had carefully considered all of the relevant facts. I shall call this group of theories *desire-based*. (*On What Matters*, §3)

Though desire-based theories come in many different forms, in what follows I will restrict my attention to the kind of desire-based theory which holds that what we have reason to do is what we would want ourselves to do if, as Parfit puts it, 'we had carefully considered all of the relevant facts', where what it means to carefully consider all of the relevant facts is in turn for those facts to have the impact that they would have in *rational deliberation* (all future references to desire-based theories should be understood as references to this sub-class).

Parfit spells out the additional commitments of desire-based theories of this kind in the following terms:

> When some desire-based theorists appeal to the desires that we would have after fully informed and *rational* deliberation, they are referring only to *procedural* rationality. According to these writers, when we are deciding what to do, we ought to think carefully about the possible outcomes of our acts, adopt aims that are easier to achieve, use our imagination, and follow certain other rules. But we are not rationally required to have any particular telic desires, or aims. We can be procedurally rational whatever we care about, or want to achieve. (*On What Matters*, §6)

There is an important concession and an equally important claim here.

The important concession is that (some) desire-based theorists do presuppose that there are rational principles governing the formation of the desires whose contents fix what we have reason to do. What we have reason to do, according to these theorists, is a function of what we would desire after the impact of information,

where the impact of that information isn't merely causal. For example, it isn't relevant to what reasons an agent has that (say) he is so constituted that he will acquire an intrinsic aversion to spiders the very first time he ever sees one. Evolutionary considerations might well afford an explanation of why all agents are so constituted, but, even if they did, that wouldn't suffice to show that it is rational for agents to respond to the perception of spiders by acquiring such an intrinsic aversion. What's relevant to what reasons an agent has, then, is the way in which the impact of information makes for *rational* changes in his desires, where this, in turn, is fixed by the 'rules' of rationality to which Parfit refers.

The equally important claim that Parfit makes in this passage is that, according to desire-based theorists who hold that what we have reason to do is a function of what we would desire that we do after informed and rational deliberation, the rules of rationality that govern the formation of desires in rational deliberation are *procedural*, by contrast with *substantive*. Though Parfit doesn't say much about what this distinction between procedural and substantive rules of rationality amounts to, he does tell us what the upshot is supposed to be of the rules all being procedural. The upshot is supposed to be that '[w]e can be procedurally rational whatever we care about, or want to achieve'. Thus, Parfit tells us, according to these desire-based theorists, 'we are not rationally required to have any particular telic desires, or aims'.[3]

As we will see, the fact that desire-based theorists are supposed to hold that all rules of rationality governing the formation of desires are procedural, by contrast with substantive, turns out to be the distinctive feature of such theories. Let's therefore attempt to classify as procedural or substantive a range of principles that various theorists have thought qualify as principles of rationality governing the formation of desires (in what follows 'RR' = 'Reason requires that'):

R1: RR (If someone 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)[4]

[3] Parfit does say a little more about what it means to be procedurally rational in the 16 April 2008 version of the manuscript. See footnote 15 below.

[4] R1 is a version of the familiar means-end principle. See Michael Smith, 'Instrumental Desires, Instrumental Rationality', *Proceedings of the Aristotelian Society, Supplementary Vol.*, 78 (2004), pp. 93–109.

R2: RR (If someone has an intrinsic desire that p, and an intrinsic desire that q, and an intrinsic desire that r, and if the objects of the desires that p and q and r cannot be distinguished from each other and from the object of the desire that s without making an arbitrary distinction, then she has an intrinsic desire that s)[5]

R3: RR (If someone has an intrinsic desire that p, then either p itself is suitably universal, or satisfying the desire that p is consistent with satisfying desires whose contents are themselves suitably universal)[6]

R4: ∃p∃q RR (If someone believes that p, then she has an intrinsic desire that q)[7]

R5: ∃p RR (People do not desire that p)[8]

R6: ∃q RR (People desire that q)[9]

There are obviously many more principles of rationality than these that theorists have posited, but these will suffice for present purposes. The question is which of these principles is procedural and which is substantive.

Principles like R1 and R2 are clearly procedural. For they simply tell us which combinations of desire and belief (in the case of R1), or which combinations of desire (in the case of R2), are rationally permissible. They also fall short of requiring us to have

[5] R2 is the minimal principle of rationality to which we need to appeal in order to rule out Future-Tuesday-Indifference. See Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984).

[6] R3 is the sort of principle Kantians think govern the formation of desires. See Christine Korsgaard, *The Sources of Normativity* (Cambridge: Cambridge University Press, 1996). Some Kantians would doubtless prefer a reformulated version of R3 that pertains to maxims or intentions. I will, however, ignore the differences between desires and intentions in what follows, as, for the most part, this distinction is irrelevant to the points I wish to discuss.

[7] As we will see, this is the sort of principle that theorists commit themselves to when they think, as Scanlon and Parfit do, that there are reasons for desiring. See T. M. Scanlon, *What We Owe to Each Other* (Cambridge, Ma.: Harvard University Press, 1998).

[8] R5 is the sort of principle to which people are committed when they think that *lacking* certain desires is constitutive of being fully rational. See, for example, Mark Johnston's response to Hume's claim that it is not irrational to prefer the destruction of the whole world to the scratching of one's finger in Mark Johnston, 'Dispositional Theories of Value', *Proceedings of the Aristotelian Society, Supplementary Vol.*, 63 (1989), pp. 139–74.

[9] R6 is the sort of principle to which people are committed when they think that *having* certain desires is constitutive of being fully rational. See, for example, Parfit's outline of various 'critical' versions of the present aim theory in *Reasons and Persons*. See also David Gauthier's comments on the way in which theorists typically argue for the claim that morality has a rational foundation (see footnote 14 below).

particular desires, as both accord with what Jay Wallace calls the 'desire-out, desire-in principle': which desires it is rational to have depends on which desires we have to begin with.[10] Desire-based theorists, as Parfit conceives of them, can therefore presumably accept principles like these. R5 and R6, by contrast, are most certainly substantive. They explicitly require us to have certain desires. They violate Wallace's desire-out, desire-in principle: which desires it is rational to have is independent of which desires we have to begin with. Desire-based theorists, as Parfit conceives of them, presumably cannot accept principles like these. But what about R3 and R4? Are these principles procedural or substantive? Can desire-based theorists accept them?

Let's start with R4. Like R1 and R2, R4 governs combinations of psychological states. It tells us that we shouldn't have certain beliefs without also having certain desires. Like R1 and R2, it thus tells us how to reason when we deliberate. To this extent, it looks somewhat like the procedural principles we've already considered. But R4 is unlike R1 and R2 in averting to beliefs and desires with particular contents rather than others: it requires us to have desires with certain contents when we have beliefs with certain other contents. It thus violates Wallace's desire-out desire-in principle: which desires it is rational for us to have depends entirely on what we believe; it is irrelevant which desires we have to begin with. Indeed, if the beliefs in question are accessible to any rational creature, then R4 might even require rational beings as such to have certain desires. An R4 style principle might therefore allow us to derive principles like R5 and R6. R4 thus also looks to be substantive. So can a desire-based theorist accept that rational deliberation is governed by a principle like R4 or not?

The answer to this question is in fact even more complicated than might initially be thought. For it is at least arguable that any desire-based theorist who accepts both that there are procedural principles of rationality like R1 and R2 and that agents are sufficiently reflective to form beliefs about the rationality of their own desires, as fixed by R1 and R2, thereby *commits* himself to at least one instance of an R4 style principle. Let me briefly explain why this is so.[11]

[10] See R. Jay Wallace, 'How to Argue about Practical Reason', *Mind*, 99 (1990), pp. 267–97.

[11] See also Michael Smith, 'Exploring the Implications of the Dispositional Theory of Value', *Philosophical Issues*, 12 (2002), pp. 329–47.

Imagine a subject who is reflective enough to form the belief that she would desire that r if she had a set of desires and beliefs that conformed to all (other) rational principles governing the formation of desires and beliefs (R1 and R2 (say)). Since there is evident dissonance in the pairing of this reflective belief with either aversion or indifference to r, it would seem to follow that she thereby commits herself to desiring that r on pain of a kind of incoherence in her psychology.[12] But, assuming now that incoherence requires the backing of a rational principle, it follows that such a desire-based theorist must grant at least the following instance of an R4 style principle governing the formation of desires:

R4$^{\text{reflective}}$: RR (If a subject believes she would desire that r if she had a set of desires and beliefs that conformed to all (other) rational principles governing the formation of desires and beliefs, then she desires that r)

To repeat, this instance of an R4 style principle follows from commonsense assumptions about the ways in which the psychological states of subjects who have the ability to reflect on the rational standing of their own psychological states must fit together. In this way it seems that we can derive substantive principles of rationality even if we initially think that all such principles are procedural.[13]

Now consider R3. The requirement that desires be suitably universal sounds like a constraint on the *form* that our desires can take. But are such formal constraints procedural or substantive? Let's say that they are procedural. But now remember that Parfit told us that it follows from a rational principle's being procedural that it falls short of requiring us to have certain desires. So does R3, which we are assuming to be procedural, require us to have certain desires? Kantians claim that this formal constraint on our desires does have a substantive upshot. Roughly speaking, they

[12] In support of the claim that there is a kind of *incoherence* in the psychology just described, note that there is a very similar kind of incoherence in the psychology of a subject who believes that he would believe that p if his beliefs conformed to all (other) rational principles, and yet who fails to believe that p.

[13] Parfit more or less concedes this point in his discussion of the relationship between *normative beliefs* and desires (*On What Matters*, §13). He doesn't tell us, though, why the upshot isn't that the class of desire-based theories, defined in the way he defines them, is simply empty.

think the only desires that are suitably universal are desires to act in ways that leave other rational beings free to lead lives of their own choosing, on condition that they in turn leave yet other rational beings free to lead lives of their own choosing, and so on. This means that Kantians, at any rate, claim that they can derive an R4 style principle, and hence R5 and R6 style principles as well, from an R3 style principle. In other words, they too think that they can derive a substantive principle from a procedural principle. However, speaking for myself, I am not sure whether the Kantians are right about this. It thus seems to me to be opaque whether R3 requires us to have particular desires. So should we suppose that desire-based theorists can accept that rational deliberation is governed by a principle like R3 or not?

The upshot of this discussion should now be clear. Parfit tells us desire-based theorists hold that the rules of rationality that govern desire formation are one and all procedural, rather than substantive, and he further tells us that it follows from this that desire-based theorists have to reject substantive principles of rationality like R5 and R6. But we have seen that principles of rationality like R3, though procedural, may, for all we know, allow us to derive principles of rationality like R4, R5 and R6; that R4, which looks to be both procedural and substantive at the same time, also allows us to derive substantive principles like R5 and R6; and that we can derive at least one instance of an R4 style principle from R1 and R2 style principles together with commonsense assumptions about the ways in which the psychological states of subjects who have the ability to reflect on the rational standing of their own psychological states must fit together. There therefore isn't a clear distinction to be drawn between theories that accept merely procedural principles of rationality and those that in addition accept substantive principles. Yet Parfit needs such a distinction if he is to rely on it in demarcating the distinctive feature of the desire-based theories of reasons for action to which he objects.

How, then, should we proceed? My own view is that we should ditch all talk of procedural versus substantive principles of rationality in the classification of theories of reasons for action. Instead we should classify desire-based theories of reasons for action more directly in terms of the principles of rationality that they take to govern desire formation, where these principles could include none or some or all of R1, R2, R3, R4, R5, and R6, and presumably many more such principles besides. This is because all such theories agree that what we have reason to do is what we would desire

that we do after informed and rational deliberation. What they disagree about is which principles of rationality govern the formation of desires and the logical relations between these principles. For example, they disagree about whether R3 governs the formation of desires and they disagree about whether certain R4 style principles follow from R3. A classification of theories of reasons for action along these lines would thus have the great virtue of focusing attention on more fundamental disagreements of this kind rather on distractions such as whether certain principles are or are not procedural or substantive.

Though this is not the place to adjudicate these more fundamental disagreements, I do want to say one thing about them as it is relevant to the discussion that follows. The ordering of the principles described above is not accidental. As we move down the list from R1 to R6 it becomes more and more controversial whether there are any principles of rationality of the relevant kind governing the formation of desires. Thus, while it is extremely plausible to suppose that our desires are governed by a principle of means-end rationality (R1); and while it is still very plausible, though perhaps a little less so, to suppose that it is irrational to have a set of desires whose contents differ in arbitrary ways (R2); and while it is still plausible, but not uncontroversially so, to suppose that our desires are subject to some sort of universalization constraint (R3); it is very controversial indeed to suppose that there are any principles of rationality governing the formation of desires like R4, R5, and R6. It should therefore come as no surprise that so much effort has gone into the sorts of attempts that Kantians and others have made to *derive* (substantive) principles of rationality governing the formation of desires like R4, R5, and R6 from (procedural) principles like R1, R2, and R3. Correspondingly, it seems that there will always be methodological reasons to prefer a desire-based theory that derives more controversial claims about the principles of rationality that govern desire-formation from less controversial. When it comes to desire-based theories of reasons for action, the weaker the better.[14]

[14] In this context it is worthwhile recalling the strategy of argument employed by David Gauthier in *Morals by Agreement* (Oxford: Clarendon Press, 1986). Gauthier attempts to argue for the conclusion that morality has a rational foundation from the premise that rational agents seek to do that which will maximally satisfy their desires: that is, based simply on the assumption that R1 is rational principle governing the formation of desires. His argument can thus be seen as an attempt to derive R4, R5, and R6 style principles simply from R1. Gauthier himself notes that it is unusual for someone who is attempting to provide morality with a rational foundation to argue in this way. The more usual strategy, he tells us, is to appeal to an 'understanding of reason' that 'already includes the moral dimension of impartiality' that he seeks to argue for (ibid., p. 6). The more usual strategy is to take it for granted that (say) 'what makes it rational to satisfy an interest does not depend on whose interest it is' (ibid., p. 7). On this more usual strategy it thus turns out to be true by definition that 'the rational person seeks to satisfy all interests' (ibid.). There is, however, an obvious problem with this more usual strategy of argument, according to Gauthier. The problem is that of showing why anyone should accept the controversial conception of rationality which it presupposes. He thus makes the methodological observation that, when attempting to argue that rationality does or does not augur in favour of acting in certain ways, we do better to argue on the basis of a conception of rationality that 'has the virtue, among conceptions, of weakness' (ibid., p. 8). Though we might doubt that Gauthier's own attempt to derive R4, R5, and R6 style principles from R1 is successful, his methodological observation is surely correct. In that same spirit, we might say that a theory of rationality that derives R4, R5, and R6 style principles from R3 also has the virtue, among conceptions of rationality, of weakness.

2. Parfit on the Nature of Value

The second main kind of theory of reasons for action, according to Parfit, is what he calls a 'value-based' theory. Before explaining what a value-based theory says about reasons for action, however, we need to remind ourselves what Parfit thinks about the nature of value itself.

The concept of value in terms of which a value-based theory of reasons for action is to be understood is what Parfit calls value 'in the *reason-involving* sense', where value in the reason-involving sense itself comes in various kinds. One kind is relativized to individuals:

> When we claim that some event would be *good for someone*, in the *reason-involving* sense, we mean that there are facts about this event that give this person self-interested reasons to want this event to occur, and that give other people altruistic reasons to want this event to occur for this person's sake.
>
> It would be in this sense good for us if we were happy, and bad for us if we were in pain, or if we suffered in other ways. (*On What Matters*, §2)

One kind of value in the reason-involving sense is thus goodness-for. For example, my own happiness is good-for-me because there is a fact about the nature of my own happiness – the way it feels and the fact it is mine – that provides me with a reason to want that

I be happy. My happiness is not good-for-you though. Rather, *your* happiness is good-for-*you*. At best, my happiness gives you what Parfit calls 'altruistic reasons' to want that I be happy, an idea that will become clearer presently.

A second kind of value is also relativized to individuals, but in a different way to goodness-for.

> We can have strong reasons to care about the well-being of certain other people, such as our close relatives and those we love. Like self-interested reasons, these altruistic reasons are both *personal* and *partial*, since they are reasons to be specially concerned about the well-being of those people who are *related to us* in certain ways. (*On What Matters*, §2)

We might call this *partial goodness in the reason-involving sense*. This is the kind of value that I would assign to the outcome of (say) my children being happy, independently of whether or not their happiness has any net effect on my own level of happiness: that is, independently of whether their happiness is good-for-me. My children's being happy is partially good in the reason-involving sense because there is a fact about the nature of my children's happiness – the way it feels and the fact that it is possessed by my children – that provides me with a reason to want that they be happy. Assignments of partial goodness in the reason-involving sense are thus always relativized to particular people as well. The happiness of my children is good_{me} and the happiness of yours is good_{you} but not vice versa. However they are not assignments of goodness-for to the people to whom they are relativized. My children's happiness is good_{me}, but it need not be good-for-me.

And there is third kind of value as well.

> We also have some reasons, I believe, to care about everyone's well-being. Such reasons are *impartial* in the sense that they are reasons to care about anyone's well-being whatever that person's relation to us.
>
> These reasons are also impartial in the different sense that these are the only reasons that we would have if our situation gave us an impartial point of view. I am using the phrase 'point of view' in something close to its literal sense, not the looser sense in which we talk of the reasons we may have from a financial, aesthetic, or other such point of view. When we think about certain possible events, our *actual* point of view is impartial. That is true when we are considering possible events that would involve or affect people who are all strangers to us. When our actual point of view is *not* impartial, we can think about possible events from an *imagined* impartial point of view. Suppose that, after some shipwreck, some rescuers could save either me or many other people who are all strangers to me. I would have strong self-interested reasons to want these rescuers to save me rather than these many strangers. But I would know that, if I were in the impartial position of some detached or uninvolved observer, I would have more reason to want the rescuers to save many people rather than saving only one.
>
> From any impartial point of view we would all have reasons, I believe, to care equally about everyone's well-being. But that is a substantive belief, not something that is implied by my definition of an impartial point of view. Some people might instead believe that, from such a point of view, we would all have reasons to care more about the well-being of certain people, such as those people who are morally best, or those who have the greatest abilities. Note next that, even when our *point of view* is impartial, that does not ensure that *we* are impartial. We might care more about the well-being of certain strangers, such as those who are more like us, or those whose faces we like. But we would have no *reasons*, I believe, to care more about the well-being of these people.
>
> We can now describe another kind of goodness. When we claim that one of two possible events would be
>
> > *better* in *the impartial reason-involving* sense, we mean that everyone would have, from an impartial point of view, stronger reasons to want this event to occur.
>
> It would be in this sense better, for example, if my imagined rescuers saved the lives of more people. It would also be in this sense better if any person, or any other *sentient* or conscious being, ceased to be in pain. This kind of goodness we can call *impersonal*. But this word may be misleading. Such goodness may be impersonal only in the sense that it is not goodness *for* particular people. Many events are made to be impersonally good by the ways in which they are good *for* one or more people, or other sentient beings. And, since everyone has reasons to want such events to occur, such impersonal goodness involves *omnipersonal* reasons. (*On What Matters*, §2)

Impartial goodness, which is different both from goodness-for and partial goodness, is the kind of goodness that we might suppose anyone's happiness has. Your happiness, or mine, or a stranger's, is impartially good because there is a fact about the nature of the happiness possessed by you, or me, or a stranger – the way it feels and the fact that it belongs to someone – that provides not just me but everyone with a reason to want that you or me or the stranger be happy. These are among the reasons Parfit calls 'altruistic reasons' to want some outcome. Assignments of impartial goodness in the reason-involving sense are thus not relativized to particular people. The happiness of anyone is simply good. It need not be good-for-me and nor need it be $good_{me}$.

As is I hope clear, this discussion of Parfit on the nature of value raises an important question. Does he really think that there are different kinds of goodness: goodness-for, partial goodness, and impartial goodness? He certainly says that goodness-for and impartial goodness are different kinds of goodness, and, as we have seen, he implies that that there is a third kind, partial goodness, as well. Or does he think that there is just one kind of goodness? There is just goodness in the reason-involving sense – this is a matter of there being some reason for wanting or caring about something – and three different kinds of reason for caring about three correspondingly different things? As we will see when we come to discuss Parfit's response to Sidgwick on the dualism of practical reason, this turns out to be a pivotal question with no obvious answer.

3. Value-Based Theories of Reasons for Action

We are now in a position to outline the second main kind of theory of reasons for action Parfit discusses.

> According to another group of theories, reasons for acting are all provided by the facts that make certain things worth doing for their own sake, or make certain outcomes worth producing or preventing. Two examples might be the facts that some act would keep a promise to someone who is dead, or would amuse someone who is bored. Most of these acts or outcomes would be good or bad for particular people, or impersonally good or bad. So I shall call these reasons *value-based.* But . . . value-based reasons derive their force, not from the goodness or badness of these acts or outcomes, but from the facts that would make them good or bad. (*On What Matters,* §3)

However, according to Parfit, for actions or outcomes to be good is a matter of there being features that provide us with reasons to want those outcomes or actions. As he puts it:

> Our reasons to have some desire are provided, I have claimed, by facts about this desire's *object,* or the event that we want. We have such reasons when the event that we want would be in some way relevantly good. We can call such reasons *object-given.* (*On What Matters,* §3)

And he further clarifies the way in which this means that value-based reasons for action depend on the facts that make actions themselves, or the outcomes of actions, worth producing in the following passage:

> On value-based theories of the kind I believe we should accept, our reasons for acting all derive their force from the facts that give us reasons to have the desires and aims that our acts are intended to fulfil or achieve. These other reasons are, in this way, more fundamental.
>
> To illustrate the kinds of claim that value-based theories make, we can next consider a few of the facts that give us reasons to have particular desires and aims. What are most important are intrinsic telic object-given reasons. These are reasons to want some possible event as an end, or for its own sake, which are provided by some of this event's intrinsic features. (*On What Matters,* §5)

According to such value-based theories, a reason for acting in a certain way thus derives its force from the fact that gives us an intrinsic telic object-given reason to want that action itself, or an outcome of the action, for its own sake.

Consider again our examples by way of illustration. The reason I have to act so as to make myself happy, which is good-for-me, derives its force from two features of my own happiness – the way it feels and the fact that it is mine – which, according to Parfit, are more fundamental reasons for me to want that I be happy for its own sake. The reason I have to act so as to make my children happy, which is partially good, derives its force from two features

of the happiness of my children – the way it feels and the fact that it belongs to my children – which are more fundamental reasons for me to want that my children be happy for its own sake. And the reason I have to act so as to make a stranger happy, which is impartially good, derives its force from two features of the happiness of the stranger – the way it feels and the fact that it belongs to someone – which are more fundamental reasons for me to want that the stranger be happy for its own sake.

Note that such value-based theorists who think that these more fundamental reasons for wanting exist thereby commit themselves to the further existence of rational principles governing the formation of desires. They commit themselves to the existence of such principles because, if there are indeed object-given reasons for desiring of the kind that they postulate, then, on the plausible assumption that these are reasons that one could follow, this entails the possibility of reasoning oneself into having the relevant desires on the basis of one's belief that those reasons obtain.

The situation is thus much the same as with reasons for believing. The fact that there are certain reasons for believing that q – the facts that p and that if p then q (say) – commits us to the existence of a rational principle governing the formation of beliefs like:

R7: RR (If someone believes that p and believes that if p then q, then she believes that q)

The existence of such reasons for believing commits us to the existence of R7 because, on the plausible assumption that we can follow these reasons, their existence entails the possibility of our reasoning ourselves into having the belief that q on the basis of our belief that these reasons obtain.

Similarly, then, the fact that there are certain reasons for desiring (say) my own happiness – the facts that happiness feels the way it does and is mine – entails the possibility of my reasoning myself into having the desire that I be happy on the basis of my belief that those reasons obtain. In other words, it commits us to the existence of a rational principle governing the formation of desires like:

$R4^{\text{good-for}}$: RR (If someone believes that a certain episode of happiness could both feel the way that happiness does and be his own, then he desires that he enjoys that episode of happiness)

But this is just an instance of an R4 style principle, that instance where p is the (egocentric) proposition that a certain episode of happiness could both feel the way that happiness does and be one's own, and q is the (egocentric) proposition that one enjoys that episode of happiness.

We are now in a position to assess Parfit's insistence that value-based theories of reasons for action are different in kind to desire-based theories. For reasons that are now evident, this claim is massively overblown. We've already seen that desire-based theories can admit the existence of principles of rationality governing the formation of desires, and we have also seen that the best way to distinguish different versions of a desire-based theory from each other is by focusing directly on what they take these rational principles to be. Different versions of a desire-based theory will claim that none or some or all of R1, R2, R3, R4, R5, and R6, and presumably other principles as well, govern the formation of desires.

Parfit's own preferred version of a value-based theory is simply a particular version of such a desire-based theory, the version according to which the principles of rationality that govern the formation of desires include at least three instances of an R4 style principle. In addition to $R4^{\text{good-for}}$, he holds that desire formation is governed by a principle corresponding to partial goodness:

$R4^{\text{partial goodness}}$: RR (If someone believes that a certain episode of happiness could both feel the way that happiness does and belong to someone with whom he has a special relationship, then he desires that that person enjoys that episode of happiness)

and that it also governed by an instance corresponding to impartial goodness:

$R4^{\text{impartial goodness}}$: RR (If someone believes that a certain episode of happiness could both feel the way that happiness does and belong to someone, then he desires that that person enjoys that episode of happiness)

In terms of our preferred taxonomy of theories of reasons for action, Parfit's value-based theory is a desire-based theory. Indeed, as we will see, it is a version of a desire-based theory that is vulnerable to a serious objection.

4. Why Parfit Prefers Value-Based Theories of Reasons for Action to Desire-Based Theories

Let's now consider Parfit's reason for preferring value-based theories of reasons for action to desire-based theories. His reason is that desire-based theories must allow, whereas value-based theories need not allow, that there are people who have no reason whatsoever to want to avoid future agony. This is an objection because, according to Parfit, it is a datum that we all have a reason to want to avoid future agony, a datum that desire-based theories cannot accommodate.

After imagining a wide variety of replies to this objection Parfit says:

> Some desire-based theorists might give a different reply. These people appeal to the desires that we would now have, or would want ourselves to have, if we had gone through some process of fully informed and *rational* deliberation. So they might claim that
>
> (F) in such cases, if we were fully rational, we would want to avoid all future agony.
>
> (F) is ambiguous. Understood in one way, (F) is a claim that would be made by any plausible value-based theory about reasons. These theories make claims about what we can call *substantive* rationality. On such theories, we all have strong reasons to have certain aims, and to be substantively rational we must have these aims. These reasons are object-given, in the sense that they are provided by the intrinsic features of what we would be trying to achieve. One such aim is avoiding future agony. If we did not want to avoid such agony, we would not be fully substantively rational, because we would be failing to respond to our strong object-given reasons to have this desire and aim.
>
> Desire-based theorists cannot make such claims. Desire-based reasons are provided, not by the intrinsic features of what we want, but by facts about what would fulfil our present telic desires. So desire-based theories imply that we have no object-given reasons to want to avoid future agony. When some desire-based theorists appeal to the desires that we would have after fully informed and *rational* deliberation, they are referring only to *procedural* rationality. According to these writers, when we are deciding what to do, we ought to think carefully about the possible outcomes of our acts, adopt aims that are easier to achieve, use our imagination, and follow certain other rules. But we are not rationally required to have any particular telic desires, or aims. We can be procedurally rational whatever we care about, or want to achieve. So desire-based theorists cannot claim that anyone who is fully rational would want to avoid all future agony. (*On What Matters*, §9)

This is the full passage, the last part of which I quoted in Section 1 above: the passage in which Parfit makes the crucial claim that what he means by a desire-based theory is a theory according to which the formation of desires is governed by principles of procedural, as opposed to substantive, rationality. In this passage Parfit seems to make the further assumption that the only way in which we can explain why people have a reason to want to avoid future agony is by appealing to something like the following R4 style principle:

> $R4^{\text{future agony}}$: RR (If someone believes that a certain future episode of agony would both feel the way that agony feels and be his own, then he is averse to the prospect of suffering that episode of agony)

and that this is a problem for desire-based theories, because desire-based theorists cannot allow that there exist substantive principles of rationality like this. They can only believe in procedural principles of rationality, procedural principles like R1, R2, and R3.

If this is Parfit's argument, then he faces many problems. For one thing, as I indicated at the end of Section 1, Parfit is simply wrong to suppose that desire-based theorists are unable to endorse any R4 style principles. Desire-based theorists as such look like they will have to accept $R4^{\text{reflective}}$, a principle which tells subjects who are sufficiently reflective to form beliefs about the rational standing of their own desires to desire that which they believe they would desire if their desires and beliefs were otherwise rational. For another – and this is the more serious problem with the argument in this passage – Parfit simply ignores the fact that there are many alternative ways in which someone could explain why we all have reason to want to avoid future agony, only some of which make a direct appeal to $R4^{\text{future agony}}$.

Consider, for example, desire-based theories of reasons for action that make a direct appeal to a principle of rationality like

R6: that is, theories according to which there is a rational requirement to be averse to present and future agony. Since this R6 style principle entails R4[future agony], R4[future agony] would still be true. On this view there would therefore still be the very reasons to want to avoid future agony that Parfit posits, but the truth of this claim would be wholly explicable in terms of the more fundamental claim that aversion to future agony is constitutive of being fully rational. Parfit doesn't even consider such a theory. Yet, since it disagrees with him about which are the most fundamental principles of rationality (recall again the taxonomic suggestion at the end of Section 1), it is a direct competitor to his own preferred version of a value-based theory which holds that R4[future agony] is fundamental.

More worrying still, desire-based theories that appeal to a universalization constraint like R3 might also secure the same result. Suppose, for example, that the Kantians are right about the way in which R3 constrains one's desires about how one is to interact with others, and suppose in addition that one's desires concerning one's present interactions with one's future self are rationally constrained in much the same way. In that case, just as the only desires concerning other people that are suitably universal, and hence rational, are those that leave those other people free to lead lives of their own choosing on condition that they leave yet others free to lead lives of their own choosing, so the only desires concerning one's future self that are suitably universal, and hence rational, are those that leave one's future self free to lead a life of its own choosing, on condition that that future self leaves its future self free to lead a life of its own choosing, and so on. On the plausible assumption that being agony free is a precondition of anyone's leading a life of his own choosing – the assumption is that the difference between mere pain and agony is that agony renders one incapable of rational choice – R4[future agony] might in this way be derivable from R3.

The real objection to what Parfit says in the passage quoted is thus that he gives no argument at all for preferring a value-based theory – that is, a theory that appeals at the most fundamental level to R4[future agony] – to one of these alternative theories. This objection is especially telling given that at least one of these alternative explanations of the reason that we all have to want to avoid future agony, the explanation based on R3, is an explanation that is plainly available to a desire-based theory even as Parfit conceives of such theories: R3 is, after all, a manifestly *procedural* principle. Moreover, given that it is more controversial to appeal

directly to an R4 style principle than it is to appeal to R3 (and here again we recall the point made at the end of Section 1), it seems that a theory that succeeds in deriving R4[future agony] from R3 is preferable to the value-based theory that Parfit himself recommends.[15]

[15] In the 16 April 2008 version of the manuscript, Parfit admits that his arguments against desire-based theories – or rather, in the terminology of the 16 April 2008 version of the manuscript, Subjectivism – do not impact upon certain versions of such theories:

> There is another kind of theory that I should briefly mention. Subjectivists, I have said, appeal only to claims about procedural rationality. But some writers, though appealing only to such claims, are much closer to Objectivists in their beliefs about what we ought rationally to want, and to do. What we ought to do, these people claim, depends on what we would want, or choose, if our desires were, in strong senses, *coherent* and *systematically justified*. On this view, if we cared about our present agony but not about our future agony, our desires would not be coherent or systematically justified. To be procedurally rational, we must care equally about avoiding agony at any time. These people also claim that, if we cared only about our own well-being, our desires would not be fully coherent or justified. To be fully procedurally rational, we must care about everyone's well-being. According to some Kantians, if we set ourselves any end or aim, we are not fully procedurally rational unless we also value the capacity of all other rational beings to set their ends, and we commit ourselves to treating others only in ways to which they could rationally consent. These people's theories are, in a way subjective, since they appeal to what we would want or choose if we were fully informed and in these demanding ways procedurally rational. But I shall use 'procedurally rational' in its ordinary thinner sense, and I shall call these people, not Subjectivists, but *Systematic Coherentists*. Like Objectivists, these Coherentists believe that we are rationally required to have certain telic desires or aims, such as a desire to avoid all future agony. That is true, Objectivists believe, because the nature of agony gives us decisive reasons to have this desire. These Coherentists defend such beliefs in a quite different way. On their view, we have no such reasons to want to avoid agony. Nor do we have such value-based object-given reasons to care about or to value other things, such as the well-being of others, or rational agency. We are rationally required to have such desires, concerns, and values, not because we have such *reasons* to have them, but because our failure to have them would make our pattern of concern incoherent, or systematically unjustified. Since such views are very different from the views that I call Subjectivist, I shall not discuss them further here. But if we have value-based object-given reasons, as I believe, we should reject Systematic Coherentism. We should claim that some things matter, in the sense that we have such reasons to care about these things. (*On What Matters*, §6)

This is a curious passage because, at least as I understand the Kantians, they insist that they too are simply appealing to the 'ordinary thinner sense' in which desires might be procedurally rational. Their claim is not that R3 is procedural in some thick sense, but rather that surprising conclusions follow from the fact that R3 is procedural in the ordinary thinner sense. Moreover, and much more importantly, Parfit misrepresents the Systematic Coherentists when he says that '[o]n their view, we have no such reasons to want to avoid future agony.' The Systematic Coherentists can agree that we have object-given reasons for wanting various things. They simply insist that the fact that we have such object-given reasons is itself explicable by a more basic fact about the rational constraints on our desires. In other words, the issue isn't whether R4[future agony] is true, but rather whether it is a fundamental truth as opposed to a truth that is derived from a more fundamental truth like R3 – or, for that matter, a version of R6. (Remember once again Gauthier's methodological observation mentioned at the end of footnote 14.)

5. Sidgwick's Dualism of Practical Reason and Parfit's Response

With the argument for the value-based theory of reasons for action in place – and from here-on I will simply assume that Parfit's preferred version of the value-based theory is correct – Parfit initially focuses on three more practical theories about our reasons for action.

According to

> *Rational Egoism*: We always have most reason to do whatever would be best for ourselves.

According to

> *Rational Impartialism*: We always have most reason to do whatever would be impartially best.

Some act of ours would be impartially best, in the reason-involving sense, if we are doing what, from an impartial point of view, everyone would have most reason to want us to do. On one view, what would be impartially best is whatever would be, on balance, best for people, by benefiting people most.

In his great, drab book *The Methods of Ethics*, Sidgwick qualifies and combines these two views. According to what Sidgwick calls

> *The Dualism of Practical Reason*: We always have most reason to do whatever would be impartially best, unless some other act would be best for ourselves. In such cases, we would have sufficient reasons to act in either way. If we knew the relevant facts, either act would be rational.

Of these three views, Sidgwick's, I believe, is the closest to the truth. According to Rational Egoists, we could not rationally act in any way that we believe would be worse for ourselves than some other possible act. That is not true. Such an act might be rational, for example, when and because we believe that this act would make things go impartially much better. I could rationally injure myself if that were the only way in which some stranger's life could be saved. According to Rational Impartialists, we could not rationally act in any way that we believe would be impartially worse than some other possible act. That is not true. Such an act might be rational, for example, when and because we believe that this act would be much better for

ourselves. I could rationally save my own life rather than saving the lives of several strangers.

On Sidgwick's view, we have both impartial and self-interested reasons for acting, but these reasons are not *comparable*. That is why, whenever one act would be impartially best but another act would be best for ourselves, we would have sufficient reasons to act in either way. (*On What Matters*, §16)

> Suppose . . . that one possible act would be impartially best, but that some other act would be best for ourselves. Impartial and self-interested reasons would here conflict. In such cases, we could ask what we had most reason to do all things considered. But this question, Sidgwick claims, would never have a helpful answer. We could never have more reason to act in either of these ways. 'Practical Reason' would be 'divided against itself', and would have nothing to say, giving us no guidance. This conclusion seemed to Sidgwick deeply unsatisfactory. (Ibid.)

But why does Sidgwick think that there is no answer to the question what we have most reason to do when self-interested and impartial reasons conflict?

According to Parfit, this is because Sidgwick accepts the 'Two Viewpoints Argument'. The Argument has three premises: (i) We assess the strength of our self-interested reasons from our own personal viewpoint ('How much do we want a certain outcome when we reflect on the effects of our action on our own well-being?'); (ii) we assess the strength of our impartial reasons from the imagined perspective of an outside observer ('How much would an outside observer want a certain outcome when they reflect on the effect of our action on everyone's well-being?'); and (iii) there is no third, neutral, viewpoint from which we can compare the strengths of these reasons. These three premises entail that the two kinds of reasons are incomparable (ibid.).

However Parfit rejects the Two Viewpoints Argument:

> This argument assumes that, when we are trying to decide what we have most reason to do, we can rationally ask this question either from our actual personal point of view, or from an imagined impartial point of view. We should reject this assumption. It is often worth asking what we would have most reason to want, or prefer, if we were in the impartial position of some outside observer. By appealing to what everyone would have

such reasons to want or prefer, we can more easily explain one important sense in which outcomes can be better or worse. But, when we are trying to decide what we have most reason to do, we ought to ask this question from our actual point of view. We should not ignore some of our actual reasons merely because we would not have these reasons if we had some other, merely imagined point of view.

Our partial and impartial reasons are, as I have claimed, comparable. Some reasons of either kind could be stronger than, or outweigh, some reasons of the other kind. And we can compare these reasons from our actual, personal point of view, whether or not this point of view is impartial. To make such comparisons, we don't need a third, neutral point of view. (Ibid.)

Parfit may well be right that we can compare partial and impartial reasons 'from our actual, personal point of view, whether or not this point of view is impartial'. But he certainly doesn't tell us why we should believe this to be so. Moreover, he seems not to notice that his own views about the nature of value make it look like this couldn't possibly be so for reasons that are remarkably similar to those invoked in the Two Viewpoints Argument. Let me explain.

Assume, just for a moment, that Parfit is a *consequentialist* in the weak sense that he thinks that we can analyse all facts about what we have reason to do in terms of facts about the value of the outcomes of the things that we can do. This is a weak sense in which someone might be a consequentialist because it makes no assumptions at all about what is of value. In particular, it makes no assumption that values are all impartial. This is important because, as we have seen, Parfit himself doesn't think that all values are impartial. He thinks that there are at least two kinds of goodness, goodness-for and impartial goodness, and he seems committed to a third kind of goodness as well, namely partial goodness. It is a weak sense in which someone might be a consequentialist for another reason too, as it allows that facts about values might themselves be constituted by more fundamental facts. This too is important, as Parfit himself thinks that facts about values are constituted by the more fundamental fact that there are features that constitute reasons for wanting. This will be important in what follows. With this weak assumption of consequentialism in place, however, together with the view that there are three kinds of goodness, we are in a position to see why Parfit's own

views about the nature of value make the conclusion that we have incommensurable reasons for action seem almost irresistible.

Suppose, for *reductio*, that my reason to save my own life, given how good for me my own life is, is *stronger than* my impartial reason to save the lives of two complete strangers given the impartial goodness of saving them. Given the truth of consequentialism, we have to be able to analyze this fact about the relative strengths of my reasons for action in terms of facts about the values of my acting in each of these ways. What this means, more specifically, is that there would have to be *more value* associated with the outcome of my saving my own life than there is in the outcome associated with my saving the lives of two complete strangers. But now remember that, in Parfit's view, there are two different kinds of goodness at stake in the two outcomes. There is the goodness-for associated with the claim that my saving my own life has value and there is the impartial goodness associated with the claim that my saving the lives of the two strangers has value. It therefore seems that all we can say is that it would be better-for-me to save my own life rather than save the lives of the two complete strangers, and that it would be impartially better to save the lives of the two complete strangers rather than save my own life. There is no third kind of goodness in terms of which we can formulate the relevant evaluative comparison. Here the similarity to the Two Viewpoints Argument should be manifest.

The upshot is that my reason to save my own life, given how good for me my own life is, cannot be *stronger* than my impartial reason to save the lives of two complete strangers, given the impartial goodness of saving them. Moreover, since similar *reductios* could be constructed to show that my reason to save my own life cannot be *weaker* than the reason to save the lives of the two complete strangers, that it cannot be *equal* in strength to the reason to save the lives of the two complete strangers, and that it cannot be *roughly equal* in strength to the reason save the lives of the two complete strangers, the only conclusion to draw is that the reasons are *incommensurable*. A conclusion much like Sidgwick's thus seems to follow from Parfit's own views about the nature of value together with the assumption of consequentialism.

It might be thought that Parfit's deeper explanation of the nature of goodness-for, partial goodness, and impartial goodness in terms of reasons for wanting somehow prevents this argument for the incommensurability of these different reasons for action from going through. He tells us, remember, that my saving my

own life has the property of being good-for-me because there are features of the outcome of my doing so that provide me with reasons of the appropriate kind to want that outcome, and that my saving the lives of two complete strangers is impartially good because there are features of the outcome of my doing that that provide me with reasons of the appropriate kind to want that outcome too. The question to ask, however, is whether these more fundamental features that constitute these values are features out of which we might construct an evaluative comparison, and a natural answer is that they are not.

There would seem to be two dimensions along which we could try to find truth-makers for such evaluative comparisons. Parfit says that we have reasons to *want* these different outcomes. But does he mean that we have reasons to want them *all-out* or *other things being equal*? If he has in mind the latter, then we could perhaps find truth-makers for the evaluative comparisons in the *strengths* of the different desires, other things being equal, that we have reasons to have concerning the different outcomes. He also says that we have *reasons* for wanting these different outcomes. But does he mean *conclusive* or *non-conclusive reasons*? Again, if he has in mind the latter, then we could perhaps find truth-makers for the evaluative comparisons in the *weight* of the reasons that we have for wanting the different outcomes. The full set of options can be represented in terms of the following matrix:

		wants	
		all-out	other things being equal
reasons for wanting	conclusive	Option 1	Option 2
	non-conclusive	Option 3	Option 4

In these terms, the important point is that if Parfit takes Option 1 – if he holds that values are grounded in *conclusive* reasons for wanting outcomes *all-out* – then, on the assumption of consequentialism, his own views about the more fundamental features that constitute the different values at stake commit him to the incommensurability of the different reasons for action thus generated. If the reasons for wanting the outcome that is impartially good and the outcome that is good-for-me are conclusive reasons to want each of these two incompatible outcomes all out, then, given

consequentialism, there is nothing to ground the claim that the reasons for action constituted by the one are stronger than the reasons for action constituted by the other.

How might Parfit respond to this argument for the incommensurability of reasons for action grounded in the features that constitute the different sorts of values? It seems to me that there are just three possible responses. One response is to accept the conclusion and revise his views about the comparability of reasons for action so as to bring his own view more in line with Sidgwick's. If he were to give this response, however, then it is important to note that he would end up with a view that is even more radical than Sidgwick's. This is because Parfit doesn't just hold that there are *two* potential sources of incommensurability in our reasons for action, but seems committed to there being a third source as well: reasons for actions are generated not just by the reasons for wanting that constitute goodness-for and impartial goodness, but also by the reasons for wanting that constitute partial goodness. Since partial goodness could conflict with both goodness-for and impartial goodness in much the same way as they conflict with each other, Parfit would therefore seem obliged to embrace not just a *dualism* of practical reason, like Sidgwick, but a *pluralism* of practical reason.

The second possibility is that Parfit might reject the consequentialist assumption that we can analyze all facts about reasons for action in terms of facts about value. If he rejects this consequentialist assumption then it is open to him to insist that, even though the different reasons for action can themselves be analyzed in terms of the values of the outcomes of the things we can do, there is a further *non-consequentialist* fact about the *strengths* of these reasons for action. This is a non-consequentialist fact because we cannot explain it in terms of any of the features that constitute the associated values. When Parfit responds to the Two Viewpoints Argument, what he says is in fact interpretable as a response along these lines. He tells us that 'partial and impartial reasons are . . . comparable . . . [a]nd that we can compare these reasons from our actual, personal point of view, whether or not this point of view is impartial'. Though, as I said earlier, Parfit doesn't explain why we should suppose that we can compare these reasons from our actual, personal point of view, it seems to me that we can now see what he might have in mind. He might be thinking that partial and impartial reasons for action have a further property, one that is not grounded in any of the features that constitute the

associated values, of having a certain strength vis-à-vis other reasons for action. In our actual situation, he might be thinking, we have the ability to detect what this further property is.

The third possible response to the argument for incommensurability was in fact foreshadowed earlier. Parfit might insist that when he says that something has value in virtue of the features that provide us with reasons to want it, he has in mind one of the other options – Option 2, Option 3, or Option 4 – and he might then construct an evaluative comparison out of the materials thus made available. For example, if he takes Option 2, then he might point out that we can sensibly ask whether, when we conjoin the conclusive reasons that I have to want the outcome of my saving my own life, other things being equal, and the conclusive reasons I have to want the outcome of my saving the lives of two strangers, other things being equal, these two reasons together provide me with conclusive reasons to want the one outcome more than the other, or for desiring the two outcomes equally strongly, or for desiring the two outcomes with a strength that is, as Ruth Chang puts it, *on a par*.[16] So long as one of these questions gets a positive answer, the values in play may turn out to be commensurable after all. For if the conclusive reasons for wanting the different outcomes together provide me with conclusive reasons to want one outcome more than the other, then we might suppose that that outcome is better; if they together provide me with conclusive reasons to want the outcomes equally strongly, then we might suppose that the two outcomes are equally good; and if they together provide me with conclusive reasons to have desires for the outcomes whose strengths are on a par, then we might suppose that the outcomes are roughly equal in value.

Note, however, that if Parfit takes this third view then he must tell us which of these three options he takes and he must explain how taking that option enables us to make evaluative comparisons. Moreover, and just as importantly, he must stop saying that goodness-for and impartial goodness are different kinds of goodness (*On What Matters*, §2). The idea that there are different kinds of goodness naturally suggests the view that the corresponding comparatives are different in kind as well – certain outcomes are *impartially better* and others are *better-for* – and then the argument for incommensurability is off and running. Parfit should undercut

[16] See Ruth Chang, 'The Possibility of Parity', *Ethics*, 112 (2002), pp. 659–88.

this argument from the outset by denying that impartial goodness and goodness-for are different kinds of goodness. He should say that there is only one kind of goodness and one corresponding comparative. If he takes (say) Option 2, then he should say that goodness is simply the property of having some feature which provides one with conclusive reasons for wanting some outcome other things being equal, and he should say that one outcome is better than another when the one has features that provide conclusive reasons for wanting it more than the other. Instead of saying that there are three kinds of goodness he should say that there are three different kinds of feature that provide one with conclusive reasons for wanting three different kinds of outcome, other things being equal (or, equivalently, that there are three different kinds of good-making feature).[17] There are the features that provide one with conclusive reasons for wanting, other things being equal, the outcome which is good-for-one; there are the features that provide one with conclusive reasons for wanting, other things being equal, the outcome which is partially good; and there are the features that provide one with conclusive reasons for wanting, other things being equal, the outcome which is impartially good.

To sum up, Parfit doesn't seem to notice that his own views about the nature of value suggest an argument for the incommensurability of reasons for action that is very similar to Sidgwick's own Two Viewpoints Argument. I have explained this argument and I have considered three possible responses to it. It would be interesting to know which of these responses Parfit would himself prefer. We can only hope that this will become clear as *On What Matters* is further refined and discussed.

[17] See also Michael Smith, 'Neutral and Relative Value after Moore' in *Ethics*, Centenary Symposium on G. E. Moore's *Principia Ethica*, 113 (2003), pp. 576–598.