Reading Parfit

Edited by
Jonathan Dancy

BLACKWELL
Publishers

bridge, Mass., 1987). For recent discussions of the importance of experiences being true or accurate, see e.g. Richard Kraut, 'Two Conceptions of Happiness', *Philosophical Review*, 88 (1979), pp. 167–97; Lynne McFall, *Happiness* (Peter Lang, New York, 1988); and John Bigelow, John Campbell and Robert Pargetter, 'Death and Well-Being', *Pacific Philosophical Quarterly*, 71 (1990), pp. 119–40.

5 See Aristotle, *Nicomachean Ethics* (*NE*), *passim* and *Rhetoric* II 12–14. See also Alasdair MacIntyre, *After Virtue* (University of Notre Dame Press, Notre Dame, Ind., 1981); Michael Slote, *Goods and Virtues* (Oxford University Press, Oxford, 1983); Annette Baier, 'Familiar Passions', unpublished paper read to the University of Cincinnati Philosophy Colloquium, 'The Concept of Emotion', March 1985; and Bigelow et al., 'Death and well-Being'.

6 My thanks are owed to Graeme Marshall here.

7 My thanks are owed to Frances Howard-Snyder here.

8 My thanks are owed to Jonathan Bennett here.

9 See Norman Care, *Living With One's Past* (Rowman & Littlefield, Lanham, Md., 1996), ch. 1.

10 My thanks are owed to Margaret Walker here. On that evaluation being of us and our acts, see my *Plural and Conflicting Values*, pp. 35–6 and 112–13; my 'Act and Agent Evaluations', *Review of Metaphysics*, 27 (1973), pp. 42–61; Norman Dahl, 'Obligation and Moral Worth: Reflections on Prichard and Kant', *Philosophical Studies*, 50 (1986), pp. 369–99; and Stephen Hudson, *Human Character and Morality* (Routledge and Kegan Paul, Boston, 1986).

11 See Korsgaard, 'Personal Identity and the Unity of Agency', and Mark Johnston, ch. 8 below. My thanks are owed to Johnston for help here and elsewhere.

12 My thanks are owed to Jonathan Bennett here.

13 On external circumstances, see John M. Cooper, 'Aristotle on the Goods of Fortune', *Philosophical Review*, 94 (1985), pp. 173–96. On whether *eudaimonia* is better understood in terms of happiness or meaningfulness and fulfillingness, see David Wiggins, 'Truth, Invention, and the Meaning of Life', *Proceedings of the British Academy*, 42 (1976), pp. 331–78; and Kraut, 'Two Conceptions of Happiness'.

14 My thanks are owed to Annette Baier here. See my 'Agent and Other: Against Ethical Universalism', *Australasian Journal of Philosophy*, 54 (1976), pp. 206–20.

15 On these last, see Michael Balint, *Thrills and Regressions* (International Universities Press, Madison, Conn., 1987).

16 On these issues, and their basis in Aristotle, see my *Plural and Conflicting Values*, pp. 63 ff.

17 My thanks are owed to Laurence Thomas here.

18 In addition to those mentioned above, my thanks are also owed to Jonathan Adler, Jeff Blustein, Chris Gowans, Elizabeth Hegeman, Lynne McFall, Liam Murphy, Ernest Wallwork and Terry Winant.

5

Parfit's P

Philip Pettit and Michael Smith

I Introduction

In *Reasons and Persons*, Derek Parfit describes two theories of rationality, the Self-interest Theory, S, and the Present-aim Theory, P. 'S and P are simply related: they are both theories about rationality' (p. 129). Parfit thinks that S represents an overwhelming orthodoxy. 'The Self-interest Theory has been believed by most people for more than two millennia' (p. 194). P is not a single theory, but rather a class of theories, and Parfit thinks that one of those versions of P which he describes as critical – CP – is the best theory. He rejects some versions of CP, but leaves a number of candidates in the field. 'We should reject the Self-interest Theory about rationality, and accept the Critical Present-aim Theory' (p. 450).

As theories of rationality, S and P say what we should do. Thus they may conflict, not just with one another and not just with other theories of rationality, but also with morality. S in particular is likely to conflict with morality. 'There are many cases where it would be better for someone if he acts wrongly. In such cases we must decide what to do. We must choose between morality and S' (p. 129).

One sort of moral theory with which S and P may conflict is the neutral sort of morality, N, which identifies a common aim that all are required to further. Parfit is concerned with the decision between theories – strictly, classes of theories – like S, P and N. We might describe these alternatives simply as theories about what we should do, specifying S and P more exactly as theories of rationality, N as a theory of morality. But it will be simpler to describe them all, in the spirit of Sidgwick (p. 129), as theories of rationality. In adopting this way of describing the alternatives, we do not beg any questions, least of all any questions against Parfit.[1]

Our primary aim in this essay is to provide a certain perspective on P

– in particular, on the group of CP-theories to which Parfit is sympathetic. We think that there are two kinds of theory that might be described as theories of rationality; one we cast as a background theory, the other as a foreground theory. This distinction will be congenial to Parfit, since he acknowledges a related contrast. Our main question is whether P-theories – particularly the preferred versions of CP – point us towards a background or a foreground theory of rationality.

The question proves to be worth pursuing, for answering it highlights various features of the theories Parfit has in mind. It turns out that all versions of P, including CP versions, can be seen as pointing us to a background theory of rationality, and that, seen in this role, they have affinities with decision theory and some variants of decision theory. It turns out on the other side that while uncritical versions of P cannot double as plausible foreground theories – in this they resemble decision theory – some critical versions can. A CP-theory can represent a plausible foreground theory, but only provided that the critical component – the component represented by C – bulks large.

Answering the question posed not only serves to highlight certain features of P-theories. It also provides a useful standpoint from which to examine Parfit's main argument against S: the argument that he describes as the 'Appeal to Full Relativity' (pp. 137–48). Looked at in the light of our discussion, that argument appears unnecessarily weak; it turns out that there is a stronger argument against S that is available to Parfit. If our primary aim is to provide a certain perspective on P, our subsidiary aim is to put Parfit under some pressure on this front.

In the next section we present the distinction between background and foreground theories of rationality. In the third section we consider P as a background theory of rationality, and then in the fourth section we look at it as a foreground theory. These sections serve to put P in perspective, in accordance with our primary goal. In the fifth and final section we pursue the subsidiary goal, using considerations raised by the earlier discussion to suggest a revision of the principal argument that Parfit presents against S.

II Background and Foreground Theories

Parfit makes a distinction between 'explanatory' and 'good' reasons. Furthermore, he makes it clear that his concern is with reasons in this second sense. 'By "reason" I shall mean "good reason"' (p. 118). The theory of rationality bears on reasons in this sense, for it is a theory about what constitutes reasons as reasons – that is, as good reasons – and about what makes some reasons better than others.

But there is a distinction to be drawn among good reasons, a distinction between two different senses in which something may be described as a good reason. This distinction is important from our viewpoint, because it generates a distinction between two kinds of theory that may each be described as a theory of rationality. The one kind we cast as a background theory, the other as a foreground theory.

The first sense of reason is that of a rational spring. A set of beliefs and desires can be a spring for the formation of a new desire or the performance of an action: say, the desire for a particular option or the performance of the corresponding action. Equally, a set of beliefs can be a spring for the formation of a new belief: say, the belief in something that follows from the contents of the other beliefs. In each case the spring may be rational; it may be a type of intentional profile that makes it rational for an agent to form the relevant sort of output, the new desire or the new belief. Other things being equal, the beliefs and desires will make the new desire for the option rational if the desires involve a pro attitude towards options with a certain property and if the beliefs involve taking the option – perhaps uniquely – to have that property.[2] Other things being equal, the beliefs will make the new belief rational if their contents – the propositions that the agent believes – provide inductive or deductive support for the content of the new belief. Where the type of intentional profile in question makes the new type of desire or the new type of belief rational in this way, it constitutes a good reason for forming that desire or belief. It is a good reason in the sense of being a rational spring for the formation of that desire or belief.

The second sense of reason is that of a rational ground. When a set of beliefs is a rational spring for the formation of a new belief, then the common presumption among philosophers is that the contents of those beliefs are rational grounds for forming the new belief. Take the beliefs that if p, then q and that p. These beliefs are a rational spring for believing that q, at least if other things are equal (other things will not be equal, for example, if the belief that q is inconsistent with something the agent already believes). Where the beliefs are a rational spring for believing that q, the (alleged) facts that if p, then q and that p, will be a rational ground for forming that belief. They will constitute a good reason for the agent to form that belief in a different sense of good reason from that of a rational spring; indeed, the difference in sense is so great that it marks a difference in *category*, as we might say. Where the rational spring consists in an intentional profile, a belief-state type, the rational ground consists in an assumed state of the world 'intended'; it consists in the way things are believed to be.

We have illustrated the notion of a rational ground with reference to the theoretical case in which beliefs lead to belief. What about the

practical case in which beliefs and desires lead to desire, and ultimately perhaps to action? In particular, what factor in a rational ground corresponds to a desire in a rational spring? Here, unlike the case with belief, there is no common presumption to provide guidance. Is the proposition that is to correspond to the desire that p the content proposition itself – the proposition that p? Is it the proposition that the agent desires that p? Is it the proposition that it is desirable to satisfy the agent's desire that p? Is it the proposition that it is desirable that p? Is it something else again? Or is there perhaps no right and wrong in the matter?

Our view, which we have defended elsewhere, is that the case of desire goes in fairly exact parallel to that of belief.[3] Just as beliefs in a rational spring correspond to certain potentially explaining and justifying propositions in a rational ground, so we think that desires in a rational spring also correspond to such propositions in a rational ground. The proposition that gives the content of a belief is what figures in the corresponding rational ground. Such a proposition is a potential explainer and justifier of having that belief in the sense that what is endorsed in assent to that proposition – the alleged fact that p – may explain, and will certainly justify, the belief that p. Our view, in parallel with this, is that the proposition that corresponds in a rational ground to a desire that figures in a rational spring is a potential explainer and justifier of having that desire. Corresponding to the desire that p will be, not the proposition that p, and not in general the proposition that one has the desire that p, but rather a proposition such that what is endorsed in assent to it makes the having of the desire suitably intelligible. With most desires the proposition will have to be that it is desirable in some way that p: it is only what is endorsed in assent to such a proposition that could suitably explain or justify the general run of desire.[4]

There are different accounts of what is endorsed in assent to a proposition like 'It is desirable that p'. They range from cognitivist accounts which take it to be a common-or-garden fact to non-cognitivist stories that represent it as a projected way of viewing the world: a soft fact, in some sense.[5] But such differences need not affect our story. The view we take is that for most desires it is such an alleged fact, whatever the ontological status the fact enjoys, which serves in a rational ground as the counterpart to the desire that p in a rational spring. Suppose then an agent acts on the desire that p together with a suitable instrumental belief: say, the belief that p can be made the case by taking a certain option O; suppose, in particular, that this desire and belief constitute a rational spring for the agent to desire, and choose, O. In that case our view suggests this: that the (alleged) fact that p is desirable – or whatever – combines with the (alleged) fact that p can be made the case by taking

option O to constitute the corresponding rational ground: the ground in view of which the agent can rationally desire O and act accordingly.

We will not repeat our earlier argument for the view that rational springs and rational grounds relate on this pattern in the practical case. We do not need to, since Parfit obviously agrees. He is prepared to countenance a sense of good reason that corresponds to a rational ground rather than a rational spring. And he is prepared to recognize that what corresponds to a desire in the rational ground is not the ascription of the desire itself, and not the content of the desire, but rather a proposition that reveals why possession of the desire is suitably intelligible. 'In most cases, someone's reason for acting is one of the features of what he wants, or one of the facts that explains and justifies his desire. Suppose that I help someone in need. My reason for helping this person is not that I want to do so, but that he needs help, or that I promised help, or something of the kind' (p. 121).

One further comment on rational grounds. As we employ the notion, a consideration X may be a rational ground for an agent's desiring something even if it happens that the agent will promote the good in question better by avoiding thoughts about X in his day-to-day deliberations. Take the good of spontaneity. We are prepared to think that the desirability of spontaneity may be a rational ground for an agent's having the desires constitutive of spontaneous behaviour even though the best way for him to promote his own spontaneity will be by avoiding spontaneity-focused deliberations in the day-to-day. It may be a rational ground, because, on reflection, in giving a rational justification for his behaviour in general – in giving a rational self-justification – the good of spontaneity may be something that he needs to take into account. It is neither here nor there that the rational thing for him to do in his more concrete deliberations is to forswear thinking about his own spontaneity.

With the distinction between rational springs and rational grounds, we are in a position to distinguish between two kinds of theory of rationality. Think of the springs as occupying a background, machine-room role, while the grounds appear in the foreground, being the considerations to which the agent actually pays attention in giving a reflective justification for his actions. The background theory of rationality will focus on good reasons in the sense of rational springs, and the foreground theory will focus on good reasons in the sense of rational grounds. In each case the theory will try to identify conditions that are necessary, and perhaps even sufficient, for an agent to have good reasons in the appropriate sense and to be, to that extent, rational; we assume that, with creatures like us, rationality requires the having of good reasons in both senses. It will tell us what is necessary, and perhaps sufficient, in the appropriate forum, the background or the foreground, for rationality.

The theory in either sense may address both theoretical and practical rationality – it may address rationality in the formation of beliefs as well as rationality in the formation of desires and in the performance of actions – but we will restrict our attention, as Parfit does, to the practical case. Our only concern will be with the background and foreground theories of practical rationality, though we shall often omit explicit mention of the practical; it is to be taken as understood.

There are a number of constraints that a background or foreground theory can recognize as conditions of rationality, and, depending on how many are introduced, the theory can be more or less demanding. It may identify a coherence condition on springs or grounds: a condition to the effect that they are internally coherent, and coherent with other states of the agent, or other things the agent posits, in a suitable way. It may prescribe a condition of reflection that should be satisfied by an agent if certain springs or grounds are to be rational ones for him to act on: say, the condition that the agent is suitably thoughtful in forming or considering the springs or grounds in question. Or it may go for a laundering condition of some kind, requiring that the springs or grounds be only of certain preferred sorts: say, that they involve only evidentially well-supported beliefs or propositions, or only certain desires or goods.[6] Theories that are demanding will tend to impose severe laundering constraints, as in the utilitarian theory of practical rationality that the only rational desire to act on in the background, and the only rational ground to take into account in the foreground, involves the maximization of happiness.

Background and foreground constraints on rationality may interact in the sense that endorsing a particular constraint in either area may commit one to endorsing a corresponding constraint in the other. To endorse the background laundering constraint according to which it is rational to desire happiness is to commit oneself to the foreground laundering constraint according to which it is rational to take happiness into account as a ground of action, and vice versa. At the limit, as is probably already evident, a background or foreground theory may be so demanding – in particular, it may involve such severe laundering constraints – that it leaves no independent questions to be resolved in the other area. Thus the background and foreground forms of the utilitarian theory of practical rationality are mutually determining in this way: if it is uniquely rational to desire the maximization of happiness, then it is uniquely rational to act on the ground of maximizing happiness, and vice versa. But it is crucial to recognize that though there are these connections between background and foreground constraints, there are some constraints in each area without any corresponding constraint in the other. This is why background and foreground theories can represent distinct areas of enquiry.

One might have expected that any foreground constraints would resolve all relevant background questions; for it is plausible to hold that an agent is rational to embrace certain beliefs and desires only if he has rational grounds for doing so: only if he has grounds laundered of certain unsuitable considerations, grounds reflectively endorsed, or grounds that cohere with one another in certain ways. But even if certain background constraints are determined by foreground constraints in this way, there may well be other background constraints to be identified. It may be, for example, that the rational agent should only form new desires – or, in the theoretical case, beliefs – that cohere in a certain way with the degrees of strength with which he holds his existing attitudes, where those degrees are determined independently of rational grounds; they are a subjective given. This, as we shall see, is the line that Bayesian decision theorists run.

What of the other possibility: that any background constraints of rationality will resolve all relevant foreground questions, rather than vice versa? Here a simple case serves to establish that this is not so. Suppose that we identify a set of constraints on the beliefs and desires that it is rational for an agent to act on. Suppose that, according to those constraints, if he has certain desires – desires that p, that q, or whatever – then it is rational for him to act on them: this is the sort of thing postulated, as we shall see, both by decision theory and by the uncritical versions of P. What does this say about what it is rational for him to do in the foreground, about what grounds it is rational for him to invoke in self-justification? It does not say enough to close all questions. For example, the rational thing for him to do in the foreground may be to look to the fact that he desires that p or that q; to look to the fact that it is desirable to satisfy such an experienced desire; or to look to the fact that it is desirable, as he sees things, that p or q or whatever.

The emerging picture is this. To commit oneself to a background theory of rationality, especially a comprehensive background theory, may be to commit oneself, at least partially, in the area of foreground theory; it may be implicitly to endorse certain foreground constraints. Equally, to commit oneself to a foreground theory of rationality, especially a comprehensive foreground theory, may be to commit oneself partially in the area of background theory. But even a fairly rich commitment in either area may not exhaust the commitments to be made in the other. There may be independent questions to be resolved on either side of the background–foreground divide.

In the next section we will focus on the background theory of practical

rationality, the theory of what makes for rational springs of desire and therefore action. In the section after that we turn to the foreground theory, the theory of rational grounds for action. In each case we will be raising the question of where P fits, and in particular where the group of CP-theories to which Parfit is sympathetic fit. We want to see how P-theories relate to well-established rivals in each area.

III P as a Background Theory

If we want to situate P among background theories of rationality, then the most useful thing to do will be first to look at the most standard theory in the area, and then try and relate other theories, P included, to that theory. So what is the most standard theory of background rationality, in particular background practical rationality? What is the orthodox account of the requirements that the potential springs of desire and action must fulfil if they are springs that it would be rational for an agent to satisfy?

The orthodox account is surely Bayesian decision theory. This theory holds that all that is required for an agent's springs of action to be rational – all that is required for an agent to be rational in serving them – is that they satisfy a certain coherence constraint. The springs of action which the decision theorist countenances are constituted by subjective probabilities and subjective utilities; these are degrees of belief and degrees of desire, where degrees are calibrated so that a rational agent will prefer something desired at a higher degree to something desired at a lower. The constraint which decision theory imposes on the subjective probabilities and subjective utilities of the rational agent is variously formulated, but in every version it comes out as a constraint of coherence.[7]

Roughly the idea is this. We focus on items of desire, items to which the agent attaches utilities, and in particular on items of desire that relate probabilistically to other desired items: thus we focus on an item like X which, as the agent sees things, will yield a desired item Y with probability $^1/_4$ and a desired item Z with probability $^3/_4$. The decision-theoretic constraint requires a rational agent to desire such an item with a degree that corresponds in a certain way to his degrees of desire for the different possible outcomes and to his associated degrees of probability. Specifically, to take the case given, it requires that the agent's utility for X be the sum of the utility he attaches to Y, multiplied by $^1/_4$, and of the utility he attaches to Z, multiplied by $^3/_4$. It requires the rational agent to conform to the rule of expected utility in the utility that he assigns to an item like X.

Let us describe an item like X as a complex object of desire: it is complex in so far as there are desired outcomes with which it is probabilistically associated. The constraint imposed by decision theory puts a requirement on the degrees of desire that a rational agent has or comes to have for such complex objects. The requirement is that the degree of desire should reflect the extent to which the object serves the agent's other desires according to his beliefs. This constraint is sometimes taken to suggest that there are simple objects of desire, and that rationality consists in instrumentally shaping one's desires for complex objects in the light of one's desires for simple ones. But the decision theorist may endorse the constraint without conceding that there are any simple objects of desire; he may work, as Richard Jeffrey does, with an atomless algebra of objects.[8] And even if he does believe that there are simple objects, he need not think that the rational agent's degrees of desire for those objects are any more primitive than his degrees of desire for more complex things.[9] This is why we describe the decision-theoretic constraint merely as a requirement of coherence among potential springs of action, a requirement of coherence among an agent's degrees of belief and desire.

Bayesian decision theory does nothing more in elucidation of rational springs than to require the coherence involved in the rule of expected utility. However, it is clearly possible to build more demanding theories of background rationality out of the theory.

Some theorists have required for rationality not only that a rational agent satisfy expected utility, but that he do so with regard to degrees of belief and desire that survive certain tests of reflection: they are considered probabilities and utilities. The basic idea here goes back at least to Sidgwick, and has been taken up by contemporary theorists like Rawls and Brandt and Gauthier.[10] Roughly, in order to be rational, we must satisfy the beliefs and desires that we would have if we were in possession of the relevant facts and could think clearly.

Other theorists have supplemented the coherence constraint of decision theory not with a constraint of reflection, but with a laundering constraint. Some of these have wanted to admit only subjective probabilities, only degrees of belief, that correspond to probabilities that are objective in some sense.[11] Others have required that an agent's degrees of desire should be ethically satisfactory, say through answering to some specified values: they have required this, if not for being rational, at least for being moral, and therefore for doing what one should. The agent must not only satisfy expected utility, he must satisfy it with regard to the utilities he ought to have, the utilities that reflect some particular values: in a phrase, he must desire according to expected value, not just according to expected utility.[12] There has not been much agreement on which

desires or preferences are unethical, as Jon Elster notes. But there has been some. 'On most accounts these would include spiteful and sadistic preferences, and arguably also the desire for positional goods, i.e. goods such that it is logically impossible for more than a few to have them.'[13]

Although we have stressed that decision theory does not have to be understood as an instrumental theory of rationality, it should be clear that the theory – that is, the theory unmodified by reflective or laundering constraints – derives from the Humean, instrumental way of thinking of rationality. We take an agent's desires as given, and we treat the requirements of rationality as requirements for the satisfaction of those desires. This approach is distinctively Humean in spirit. Of course, the expected utility version of what it is to be rational, what it is to satisfy desires, is much more complex than anything which Hume envisaged, but that should not surprise us, given the use it makes of quasi-mathematical notions that do not figure in Hume. The greater complexity means that the theory places requirements on a rational agent's desires of a kind that Hume sometimes seems to rule out. Thus it requires that an agent's desires satisfy various subsidiary conditions of coherence; for example, it requires that an agent's preferences not be intransitive: if he prefers *A* to *B* and *B* to *C*, then he cannot prefer *C* to *A*. But this sort of requirement is not imposed wilfully. As things turn out, it makes no sense to require an agent to desire and act according to the rule of expected utility if he has an intransitive preference structure.

Not only is decision theory – that is, unconstrained decision theory – Humean in spirit, however. It also represents the most sophisticated attempt to give expression to the orthodox Humean notion of what practical rationality – background practical rationality – consists in. For that reason it stands out as the background theory of rationality which most contemporary philosophers, economists and social theorists would identify as orthodoxy. It explicates what John Rawls describes as the 'standard' and 'familiar' concept of rationality.[14] As we have seen, theorists of rationality – and especially morality – often constrain decision theory with different sets of reflective and laundering constraints. But the point to notice is that unmodified decision theory is for so many of them the natural place to start. It is the agreed rest position, the position from which other destinations are most easily reached.

With decision theory in place, we may now ask where P stands in relation to it. In its initial formulation – one amended in critical versions, as we shall see – P says that what an agent has most reason to do is whatever would best fulfil his present desires (p. 117). This statement brings out a striking contrast between P – in all versions – and decision theory. Decision theory says that what it is rational for an agent to do – what complex object it is rational to desire – is whatever best serves his desires *according to his beliefs*. P says that what he has most reason to do is whatever best serves his desires *in fact*. This might mean: whatever best serves his desires, given suitably objective probabilities as to how things will turn out. Parfit, however, takes it to mean: whatever actually best serves his desires, whatever will turn out to serve them best, however improbable its turning out that way is at the time of action. We can put this initial contrast between decision theory and P by saying that P objectifies decision theory.

This difference between decision theory and P-theories need not mark any disagreement. The question of what it is rational to do can be treated in either of two ways, as a question about what is subjectively rational or as a question about what is objectively so. The subjective question concerns what it is rational for an agent to do in the light of his beliefs, where the answer will vary with varying beliefs. The objective question bears on what it is rational for the agent to do in a sense – assuming there *is* a legitimate sense – in which the answer is not supposed to vary with a variation in the agent's beliefs; it is the question, as Parfit likes to phrase it, of what an agent has most reason to do. A complete theory of rationality, as Parfit agrees, will address both questions – though it may not treat them as of equal legitimacy or importance (pp. 25, 120, 153). This is unsurprising, since an answer to one will tend to suggest a line on the other. But though a complete theory will address both questions, it is common in discussions of rationality to focus on one or the other.

The difference between decision theory and P-theories is that decision theory is designed to answer the question about subjective rationality, whereas P-theories are generally formulated by Parfit to answer the question about objective rationality; they are presented as theories about what an agent has most reason to do. This is not to say that Parfit thinks that P-theories can only deal with the objective question. He opposes such theories to the Self-interest Theory, S, and just as he thinks that S has an answer to the subjective as well as the objective question, so he presumably thinks that P theories can be adapted to provide an answer to the subjective issue (p. 8). Indeed, the indications are that if they were adapted to cope with the subjective question, Parfit's P-theories would look very like decision theory; his discussion of S in this role suggests that they would make use of the decision-theoretic notion of expectation.

When we say that P objectifies decision theory, then, we do not say that it diverges from it. All we mean is that it represents a counterpart of decision theory – one that mentions desires but ignores beliefs – which is suited to dealing with the objective as distinct from the subjective question about rationality. Is P – strictly, the unqualified version of P – the only objective counterpart of decision theory? No, as already implicitly noted. A decision theorist of subjective rationality says that an agent

should do whatever best serves his desires according to his beliefs. Asked to address the question about objective rationality – if he admits its legitimacy – he might say, not that an agent should do whatever actually best serves his desires, but that he should do whatever best serves his desires according to some appropriately objective probabilities at the time of action.[15] Such a theory might equally well claim to objectify decision theory.

Parfit introduces two amendments of the initially formulated doctrine, the doctrine that just objectifies decision theory. First, he assumes that the desires that a rational agent has most reason to satisfy – the desires which it is objectively rational for him to fulfil – are those that would survive a certain reflection on the agent's part: desires that he would have 'if he knew the relevant facts, was thinking clearly, and was free from distorting influences' (p. 118). The condition of reflection introduced here is the familiar sort discussed earlier in relation to decision theory. Notice in particular that it is reductively specifiable by reference to what are the facts, what are relevant unclarities, and what are typical distorting influences; reflection will have occurred whenever the agent operates in knowledge of the facts and in the absence of such unclarity or distortion. Parfit does not presuppose an independent notion of rational desire such that we could then characterize the condition where reflection has occurred, in a non-reductive way, as whatever condition is necessary for the formation of rational desire: whatever condition it takes to produce such desire. Parfit discusses only cases where the reflective condition is met, and we shall generally take the condition as given in what follows (p. 120).

So far the sort of P-theory envisaged comes to an objectified and reflective version of decision theory. Parfit also introduces a second qualification to the variants of P that he countenances, a qualification that amounts to a laundering constraint. He is prepared to countenance only critical versions of P: only CP-theories. P goes critical in virtue of two distinct sorts of constraint over and beyond the reflective condition (p. 119). One is a constraint already assumed, as we have seen, in decision theory: that the agent's desires overall satisfy coherence conditions like that of not involving intransitive preferences. This we may ignore. The other is a constraint on the content of individual desires and sets of desires.

Parfit requires that individual desires are not irrational in any intrinsic way, that they are not open to rational criticism. He cites as compelling examples of irrational desires those desires that would discriminate between pleasures and pains in some arbitrary way: desires like the desire to avoid pain except on Tuesdays (pp. 124–5). Parfit is also open to the possibility that some desires are rationally required, and that a set of

desires may be irrational in content, through failing to contain such desires (pp. 119, 121–2). Thus the group of CP-theories that he is prepared to countenance covers a broad range, stretching from the weak sort that would just outlaw irrational desires like the ones mentioned to the strong sort that would represent some desires as rationally required.

Can we say more about the theories that belong to Parfit's preferred group? All theories that address the question of what it is objectively rational for an agent to do can be presented in CP terms; the CP structure is one of which '*every* possible theory about rationality is one version' (p. 194). But Parfit is hostile to some CP-theories. He rejects the CP version of S, CPS, according to which what a rational agent should do is whatever actually best serves the desire that is alleged to be supremely rational: namely, that things go as well as possible for an agent over the course of his life (p. 131). Equally, of course, Parfit rejects the null version of CP – null because the critical element vanishes – which would deny that any desires are rationally impermissible or mandatory (p. 194). Otherwise he is well disposed to CP-theories. Thus he is open to the idea that the desire that things go as well as possible for an agent over the whole of his life may be rationally required, even if not supremely rational (p. 135). And he is open also to the idea that the desire to further certain moral ends may be rationally required, and even indeed that it may be supremely rational (pp. 121–2, 133, 452). This theory, CPM, is the CP version of N, as CPS is the CP version of S.

We have situated Parfit's P-theories in the context of other background theories of rationality, presenting them in relation to decision theory and well-known variants of decision theory: those that would impose extra reflective and laundering constraints. All P-theories differ from decision theory in being formulated as objectified theories, as theories that are addressed to the question about objective rather than subjective rationality. And CP-theories differ from unmodified decision theory – though not, of course, from the constrained variants – in being subjected also to reflective and laundering constraints.

This presentation of P-theories should be useful in providing a perspective on them. It does not jar with anything that Parfit himself says, except in one minor respect. It suggests that Parfit overstates the novelty of his own theory and the standing of S. 'It has been assumed, for more than two millennia', Parfit says, 'that it is irrational for anyone to do what he knows will be worse for himself' (p. 130). S is, he tells us, 'the verdict of recorded history', and he therefore worries that we will find 'absurdly rash' his rejection of S in favour of P, and in particular CP (p. 194). But these claims about S are overstated, and this anxiety is groundless.

While there have certainly been adherents of S for more than two millennia, over the last 200 years it has been even more widely assumed

that it is not irrational for an agent to do what he knows will be worse for himself. ' 'Tis not contrary to reason', as Hume put it, 'for me to chuse my total ruin, to prevent the least uneasiness of an Indian or person wholly unknown to me.'[16] Hume transformed our way of thinking about practical rationality, and that transformation has culminated, over the past half-century, in the development and widespread acceptance of decision theory as a formal model of practical rationality. It is decision theory, unconstrained by the requirement that the rational agent's desires answer to the specific value of furthering his own interests, that now enjoys the status Parfit claims for S over the last two millennia.

P-theories are closely related to decision theory, as we have seen. Unqualified versions of P objectify decision theory, while the critical versions, in particular those that Parfit himself favours, also impose certain reflective and laundering constraints. Thus the endorsement of a CP-theory in preference to S – or even in preference to unconstrained decision theory – does not represent an 'absurdly rash' move. On the contrary, it is likely to seem a natural and reasonable initiative.

IV P as a Foreground Theory

Our discussion in the last section shows that it is perfectly natural to take P-theories as background theories of rationality. There is more difficulty, however, in taking them as foreground theories, as we shall now see. It turns out that the unqualified version of P, and some of the critical versions too, cannot plausibly be taken in a foreground role. This is not a criticism of those theories, but an interesting fact about them. It indicates that they should be seen as theories which address only background questions of rationality. In this limitation, as we shall see, these theories resemble decision theory.

The background theory of practical rationality is concerned with what makes for rational springs of action. The foreground theory addresses the parallel question of what makes for rational grounds of action. We saw that a background theory may look to constraints of coherence, reflection or laundering in formulating the requirements of rationality. Equally, a foreground theory might look to constraints of such kinds. It might say that what is required for certain grounds to be rational is that they cohere with one another in a certain way; that they are grounds that the agent would endorse on reflection; that they exclude certain unsuitable sorts of considerations; or a mixture of such things.

The two sorts of constraints that are actually most invoked in the foreground theory of rationality are coherence and laundering constraints. The best-known constraint of coherence has it that if a consideration gives an agent a good foreground reason – a rational ground – for doing something, and if it mentions a particular individual, time or place, then considerations that differ at most in the particular mentioned must give rational grounds to any individual in the position of the agent. It would be incoherent to acknowledge the one consideration as a good reason without acknowledging the others as good reasons. This is the constraint of universalizability. It means that if I am given a good reason for doing something by the fact that a particular individual is in need, then anyone in my situation would be given a good reason for doing that sort of action by the fact that any relevantly similar individual was in need. And it also means, to take a slightly more complex example, that if I am given a good reason for doing something by the fact that it will help my child, then anyone in my situation would be given a good reason for that sort of action by the fact that it would help his or her child.

Theories of foreground rationality often also introduce laundering constraints on the considerations that may rationally justify an action from an agent's point of view. Thus a theory might prescribe that the fact that an action will cause one pain constitutes a good reason *pro tanto* for not doing it, or that the fact that an action will cause one pleasure always constitutes a good *pro tanto* reason for doing it. Again, a theory might prescribe that the only consideration that can rationally justify an action to an agent is the fact that so acting will maximize happiness generally. And so on through other salient examples.

It should be clear that a satisfactory theory of background rationality will not necessarily double as a satisfactory theory of foreground rationality. That is to say, a formula as to what one has good or most reason to do may constitute a satisfactory theory as it applies to rational springs without constituting a satisfactory theory, or even a half-sensible theory, as it applies to rational grounds. Suppose we endorse decision theory: we think that any set of subjective probabilities and utilities will constitute rational springs for an agent to act on, provided they satisfy the coherence constraint associated with the agent's maximizing expected utility; moreover, we think that these are the only rational springs there are. This does not mean that we will take decision theory to provide also a satisfactory theory of foreground rationality: a theory that closes the foreground questions which, as a background theory, it leaves open. These will include questions to do with whether the rational agent should focus on considerations about which desires he has, considerations about the desirability of satisfying those desires, or considerations about the desirability of the things he desires.

Our own view is that decision theory sticks entirely to background matters, and has nothing to say on such foreground questions. Someone who thinks that decision theory serves also in a foreground role – the

business school enthusiast perhaps – will take a very different view. He will hold that the decision-theoretic formula offers advice on how to deliberate, suggesting that the agent should consult his own degrees of belief and desire in deciding how to act; he should consult these, rather than the alleged facts that support them, the facts about what is so and about what is desirable. We think that what is proposed here has little merit or sense.

As a foreground theory, decision theory would say that any suitable set of subjective probabilities and utilities will constitute rational grounds for an agent to take into account in self-justification, and indeed that there are no other sorts of rational grounds available. It would counsel the agent to consider the state of his beliefs and desires with a view to determining in every choice the option which best serves those beliefs according to those desires, the option which maximizes expected utility. But it would be crazy to prescribe that an agent should deliberate only from considerations as to what he believes and desires, as distinct from considerations as to what is the case or what is desirable. We cannot seriously entertain the possibility that the rational agent should not consider the things he believes – that p, that if p then q, and so on – in deriving and justifying new beliefs, but should rather consider the fact that he believes that p, believes that if p then q, and the like. Neither can we countenance the possibility that he should not consider the factors that serve to justify and explain his desires – that an option will help a friend, that it will make him famous, or whatever – but should rather focus on the desires themselves.[17] We need not argue the point, since Parfit would obviously agree with us. He is explicit, as we have seen, on the case of desire: 'my reason is not my desire but the respect in which what I am doing is worth doing, or the respect in which my aim is *desirable* – worth desiring' (p. 121).

We have seen that Parfit's P-theories represent fairly reasonable proposals on matters of background rationality, being objectified and, in the case of CP-theories, reflective and laundered versions of decision theory. The question now is what a P-theory would amount to as a theory of foreground rationality, a theory as to the grounds which a rational agent will take into account for choice. The question in particular is whether it would make for a foreground theory of a more plausible kind than decision theory or whether, like decision theory, it is best seen as a purely background theory.

The unqualified version of P, a version rejected by Parfit himself, says that the rational thing for an agent to do in any situation is what best serves his desires in fact. It will be clear that this version of P cannot serve as a foreground theory any more successfully than decision theory. Apply the P-formula to resolving foreground rather than just background

questions, and it would have the rational agent restricted to a consideration of the state of his desires. The unqualified version of P is subject to the complaints just made about decision theory as a foreground theory, complaints which we assume that Parfit would support. Thus this version of P should be seen only as a background theory of rationality.

At the other extreme from the unqualified version of P are the critical versions which qualify P to the extent of taking a particular sort of desire to be supremely rational. We are thinking here of CPS, which confers this privilege on the desire that one's life go as well as possible, and CPM, which gives a similar status to certain moral concerns: say, the utilitarian concern with overall happiness. Such theories can serve as reasonable theories of foreground rationality – certainly as theories less mad than decision theory – because they can be taken to prescribe as grounds for the rational agent to consider, not matters to do with the satisfaction of a desire that the agent happens to have, but rather matters concerning what would be desired if the agent were fully rational: the rational agent's own well-being or the happiness of sentient creatures overall. It is not crazy to claim that the only rational ground for choosing something is that it is supremely rational to desire it.

What now of Parfit's preferred group of CP-theories? These, as we have seen, will include theories as strong as CPM, theories that represent the desire for some moral end as rationally required and supremely rational. There will be no difficulty, as we have just seen, in taking such theories in a foreground as well as a background role; there will be no problem in seeing them, not just as telling us the desires it is rational to satisfy – only those that serve the supremely rational desire – but also as offering counsel on the grounds it is rational to invoke in deliberative self-justification.

But Parfit's preferred theories also include theories that are very weak in the critical dimension: theories that rule out only desires that are irrational in certain conspicuous ways – say because they distinguish arbitrarily between different instances of a certain sort of pleasure or pain. What is the position going to be with such theories? Will they be capable of doubling in a foreground role? Or will they be better taken, like decision theory and the unqualified version of P, as theories of a purely background kind: theories that do not address, or that address only in part, questions about the grounds that it is rational to invoke in deliberation?

What sorts of grounds would such a weakly critical CP-theory prescribe that the rational agent should consult in deliberation? Because it insists that certain desires and patterns of desire are intrinsically irrational, it will certainly rule out certain considerations from being taken into account by a rational agent. The rational agent will not be moved, at

bottom, by the thought that pains are undesirable except on a Tuesday; he will not be moved by judgements of desirability that fit a pattern whereby X is more desirable than Y, Y than Z, and Z than X; and so on. But for the rest, it seems, such a minimally critical version of CP, interpreted as a formula for resolving foreground questions, would simply tell the rational agent to look to the state of his own present desires, provided they meet certain minimal conditions of reflection, in determining what to do. And in that case it follows that the theory would run into the same troubles as decision theory if it were taken as a theory of foreground rationality. It would prescribe as grounds that are uniquely appropriate for rational choice considerations that Parfit himself tells us do not in general constitute such grounds. For these grounds do not concern the worth or desirability of actions.[18] We should take a theory of this kind to be a theory that addresses background, not foreground, questions about rationality.

As against this line of argument someone may say that every foreground theory of rationality must acknowledge that there are some desires which an agent will have without reasons – hankerings, hungers and the like – which will provide *pro tanto* foreground reasons, of themselves, for action: that is, for acting so as to satisfy them. So what is supposed to be so counter-intuitive about the foreground proposals that a weakly critical CP-theory would put forward?

There are two points to make. First, the sorts of desires quoted in analogy are unusual, having a dual aspect as producers of action and as phenomenological yearnings, and it is going to be strange if a theory – a weakly critical CP-theory – treats other kinds of desires as of the same sort: say, if it treats in this way my desire to write a novel, be kind to friends, or become famous. Second, even it were fair to treat other kinds of desire like these phenomenological inclinations, this would not support the conclusion that such desires offer, of themselves, foreground reasons to act. If I act on a phenomenological desire, my foreground reason must be that it is desirable in some way to satisfy it, not just that I have the desire; the latter reason could be as much a reason to get rid of the desire, say by therapy or by resort to a cold shower, as it would be a reason to satisfy it.[19] The first of our two points shows that the objection is based on a strained analogy, the second that it is based on an analogy which fails to deliver the required result.

We saw in the last section that any P-theory – in particular, any of the CP-theories countenanced by Parfit – can pass muster as a background theory of rationality. In this section we have seen that some strongly critical CP-theories, including CPS and the sort of CPM which Parfit keeps in his preferred group, can reasonably double as foreground theories too. But we have also seen that the unqualified version of P, and less

strongly critical CP-theories, cannot reasonably be taken in this role; they cannot reasonably be taken as offering advice on rational grounds for deliberation. Like decision theory, such theories are better taken as doctrines addressed solely to questions of background rationality.

V Parfit's Case against S

Our argument so far serves to put P-theories – in particular, the CP-theories preferred by Parfit – in some perspective. We hope that it throws light on their relations with other theories of rationality and on their capacity to answer the different sorts of questions that come up in the background and foreground areas. The argument does not support any criticism of Parfit, except in the suggestion that he overstates the standing of S and the novelty of his own proposal.

In this final section, however, we do mean to offer a challenge to Parfit. We consider one of his main arguments against S – the Appeal to Full Relativity (pp. 137–48) – and find that his failure to be explicit about the distinction between background and foreground theories of rationality leads him astray. In that argument the unqualified version of P is evidently treated as a foreground theory, and as a foreground theory that is plausible in its own right. For reasons rehearsed in the last section, such an argument should convince no one, least of all Parfit himself. But though the argument as presented does not succeed, our discussion reveals that a closely related, and indeed stronger, argument remains available. The criticism we offer, therefore, is constructive in effect, and Parfit may find it congenial.

Parfit introduces his argument[20] with these remarks:

> Sidgwick's moral theory requires what he calls Rational Benevolence. On this theory, an agent may not give a special status either to himself or to the present. In requiring both personal and temporal neutrality, this theory is *pure*. Another pure theory is the Present-aim Theory, which rejects the requirements both of personal and of temporal neutrality. The Self-interest Theory is not pure. It is a *hybrid* theory. S rejects the requirement of personal neutrality, but requires temporal neutrality. S allows the agent to single out himself, but insists that he may not single out the time of acting. He must not give special weight to what he *now* wants or values. He must give equal weight to all the parts of his life, or to what he wants or values at all times.
>
> Sidgwick may have seen that, as a hybrid, S can be charged with a kind of inconsistency. If the agent has a special status, why deny this status to the time of acting? We can object to S that it is *incompletely relative*. (p. 140)

Parfit's argument against S is a pincer argument that S is incoherent in recommending a partiality to self over others but an impartiality as between present and future times. Parfit endorses a principle of full relativity or partiality: this is the Appeal mentioned. According to this principle, reasons – in particular, the allegedly compelling reasons countenanced by S – should be fully relative if they can be relative at all: they should be fully relativized to persons and times if they can be relativized to either (pp. 140–1). His defence of the principle is that any grounds for going relative in one way will be grounds for going relative in the other. If we go agent-relative in the theory of what it is rational to do, arguing that the question before me as an agent is what there is most reason for *me* to do, we should go time-relative also, on the grounds that equally the question that I face is what it is rational for me to do *now*: I should say that what it is rational for me to do is whatever will promote the good for me now (pp. 142–3). Thus, the pincer argument says, the defender of S should give up his theory in favour of either a fully relative theory or a fully neutral one; he should say that what it is rational for an agent to do is to promote the good generally over all people and times or to promote the good for himself now. He should not treat 'I' and 'now' in less than an even-handed way (p. 148).

This argument is directed principally against S as a foreground theory, arguing that S offends against a certain coherence constraint. The argument is that we cannot rationally justify our conduct to ourselves by appeal to considerations that are only incompletely relative, such as the consideration that this option will be beneficial for me over a range of times. The rational agent who considers conforming to S – the agent who thinks that there are rationally compelling considerations of the kind cited by S – is enjoined by the pincer argument to prefer to countenance considerations that are fully neutral or fully relative. Thus, on Parfit's account, the argument produces an instability result for the Self-interest Theory. It means that the theory is a half-way house between foreground theories recommending that we act on fully relative and fully neutral considerations. We should cease to look for a theory of compelling reasons altogether, or we should reject S in favour of one of the extreme positions.

But though the argument is directed principally against S in a foreground role, it tells against S in a background role as well. If we are required rationally to abjure any foreground theory of compelling considerations, such as S, or to reject S as a foreground theory in favour of a theory that recognizes fully relative or fully neutral considerations, then in the theory of background rationality we must forswear any demanding theory like S – any theory requiring a certain desire – or we must reject S in favour of theories that launder desires uniformly: theories which require rational desires to have either fully relative or fully neutral contents. We cannot endorse a theory like S which casts as supremely rational a desire that is intertemporally neutral but interpersonally relative.

The account of Parfit's argument offered so far is silent on the question of what constitutes the good which S says the agent should promote for himself over his life as a whole. S says that an agent should act so that his life goes as well as possible, but in our account of this argument against S we have said nothing about what it is for a life to go as well as possible. S may be interpreted in any of a number of different ways, depending on what the good for a person is taken to be. Thus S will vary in its concrete significance, Parfit says, as the good is equated with pleasure or the fulfilment of the agent's desires or the realization of an objective benefit like knowledge (p. 4).

When Parfit develops his pincer argument against S, he does so under a particular interpretation of the good: namely, an interpretation which equates the good for an agent with the fulfilment of his desires (p. 137). On this construal, S says that the only considerations that can rationally justify his conduct to an agent are those that concern the satisfaction of his desires over his lifetime. The pincer argument alleges that he should, rather, act on fully relative or fully neutral considerations: considerations to do with what will satisfy his desires now or considerations to do with what will satisfy people's desires generally.

What doctrine prescribes that the rational agent should act on considerations about what will satisfy people's desires generally? Rational Benevolence, in something like Sidgwick's sense: specifically, utilitarianism in a desire-centred form. And what doctrine prescribes that the rational agent should act on considerations to do with what will satisfy his own desires now? The Present-aim Theory, Parfit tells us (p. 140; see also p. 135). And here we see the promised dénouement. In pressing his pincer strategy against S, Parfit commits himself to the view that, of the two foreground theories of rationality that are to be preferred to the desire-fulfilment version of S, the fully relativized alternative is P in the unqualified form that says that the rational agent should do whatever would best satisfy his present desires.

But this has to be a mistake on Parfit's part, since we know he thinks that considerations regarding their desires do not in general give people good foreground reasons to act: that they should act instead on considerations that serve to explain and justify their desires, considerations as to the worth of what they desire (p. 121). We think that he could never have made this mistake if he had applied the background–foreground

distinction to his discussion of good reasons and of theories like S and P. It is a pity that though he appears to be committed to such a divide, he did not give more explicit attention to it in his treatment of these topics.

What is the effect of admitting, as we think Parfit should admit, that one of the pure alternatives with which he seeks to destabilize the desire-fulfilment version of S is an alternative that has to be rejected out of hand? Surprisingly, the effect is to strengthen the case against this form of S. Parfit's destabilizing argument is that anyone who adopts a theory of compelling reasons like S has a reason to prefer a fully relative theory or a fully neutral one. But if the fully relative theory is not a real option, then the argument would seem to establish that anyone who adopts S has a reason to prefer a fully neutral theory. We might cast the argument as follows, in the form of a *reductio*. The relativization countenanced in adopting the desire-fulfilment version of S leads in consistency to full relativization; but full relativization is in this case quite objectionable, involving the adoption of the unqualified version of P as a foreground theory; so relativization of the sort involved in this form of S should not be tolerated. So stated, the argument resembles the case made by Thomas Nagel in *The Possibility of Altruism* against the relativization of foreground reasons.[21] If we are to countenance compelling reasons of the kind alleged by the desire-fulfilment version of S, then, we should countenance only neutral reasons as compelling.

Someone may say that while we have shown that Parfit has a *reductio* strategy available against S, with the good interpreted as desire-fulfilment, we have not shown that he has such a strategy available under other interpretations of the good. It is true that we have not shown this, but it turns out that something close to the more general result can be established. This is surprising, since it makes it even less explicable why Parfit should have thought that he had only a destabilizing argument available against S.

The other interpretations of the good for a person that Parfit cites equate it with pleasure or with a more objective good like knowledge (p. 4). Suppose, then, that the good is taken as pleasure or knowledge. The Self-interest Theory will prescribe that the rational agent should take considerations about the promotion of his own pleasure or knowledge over his lifetime as supremely compelling: as considerations that trump everything else. The pincer argument says that, on the contrary, if such trumping considerations are countenanced, then the only coherent recommendations are either that the agent should act on considerations about the promotion of pleasure or knowledge generally – the fully neutral position – or that he should act on considerations about the promotion of his pleasure or knowledge now – the fully relative position. Thus it appears that the mid-way position S is unstable.

But in this case, as in the argument with the good taken as desire-fulfilment, it appears that Parfit has the resources available to do more than just destabilize S. He himself points out that if we apply relativization to the hedonistic version of S, so that the happiness of me now is what counts, we are led to an 'absurd' view (p. 142; see also p. 135). And in that same context he does not even mention knowledge – or any other such good – as the sort of thing that might plausibly be held to be in the interest of me-now; he would therefore presumably regard the theory that would focus on present knowledge rather than present pleasure as equally, if not more, absurd. But if the pincer argument shows that a neutral theorist who goes to S ought in consistency to go to a position that is rated absurd, then the argument does more than destabilize S; it reduces S to absurdity and, among theories that countenance compelling considerations, establishes the unique superiority of the neutral position.

The lesson of these reflections is that not only does the introduction of the background–foreground distinction help in situating P-theories relative to decision theory and other theories of rationality. It also helps us to see that Parfit goes astray in the course of one of his main arguments against S: specifically, in thinking that he has only a destabilizing strategy available against S, when a *reductio* strategy is accessible from the very considerations he musters. This charge may not be uncongenial to him. Not only is S unstable, as he alleges. It involves a sort of relativization that leads to near-absurdity.[22]

Notes

1 True, we abstract away from the distinction between what is subjectively and what is objectively rational, which is important in Parfit's work, but we shall lift that abstraction later.

2 Donald Davidson, 'Actions, Reasons and Causes', in *Essays on Actions and Events* (Clarendon Press, Oxford, 1980), pp. 3–19.

3 Philip Pettit and Michael Smith, 'Backgrounding Desire', *Philosophical Review*, 99 (1990), pp. 565–92. See too a paper written five years after 'Parfit's P': viz. 'Freedom in Belief and Desire', *Journal of Philosophy*, 93 (1996).

4 Michael Smith, 'Valuing: Desiring or Believing?', in *Reductionism, Explanation and Realism*, ed. David Charles and Kathleen Lennon (Oxford University Press, Oxford, 1992), pp. 323–60. For a later, fuller treatment, see *The Moral Problem* (Blackwell, Oxford, 1994), esp. ch. 5. Why do we say 'with most desires'? Because with some desires – visitations like hungers and hankerings – the rational ground for acting, if there is one, will be that it is desirable to satisfy such urges. See the discussion at the end of section IV.

5 See e.g. Simon Blackburn, 'How to Be an Ethical Anti-realist', in his *Essays in Quasi-Realism* (Oxford University Press, Oxford, 1993), pp. 166–81.

6 Robert Goodin, 'Laundering Preferences', in *Foundations of Social Choice Theory*, ed. J. Elster and A. Hylland (Cambridge University Press, Cambridge, 1986).
7 Ellery Eells, *Rational Decision and Causality* (Cambridge University Press, Cambridge, 1982).
8 Richard C. Jeffrey, *The Logic of Decision*, 2nd edn (University of Chicago Press, Chicago, 1983).
9 Philip Pettit, 'Decision Theory and Folk Psychology', in *Essays in the Foundations of Decision Theory*, ed. M. Bacharach and S. Hurley (Blackwell, Oxford, 1991), pp. 147–75. For a later, fuller treatment, see Pettit, *The Common Mind: An Essay on Psychology, Society and Politics* (Oxford University Press, New York, 1993), esp. ch. 5.
10 See Henry Sidgwick, *The Methods of Ethics* (Dover, New York, 1966), bk 1, ch. 9; John Rawls, *A Theory of Justice* (Oxford University Press, Oxford, 1971), sect. 64; Richard Brandt, *A Theory of the Good and the Right* (Oxford University Press, Oxford, 1979), ch. 6; and David Gauthier, *Morals by Agreement* (Oxford University Press, Oxford, 1986), ch. 2.
11 See e.g. D. H. Mellor, 'Objective Decision-Making', *Social Theory and Practice*, 9 (1983), pp. 289–310.
12 See e.g. Frank Jackson, 'A Probabilistic Approach to Moral Responsibility', in *Proceedings of the 7th International Congress of Logic, Methodology and Philosophy of Science* (1983), ed. R. Barcan Marcus et al. (North-Holland, Amsterdam, 1986), pp. 351–66; Mellor, 'Objective Decision-Making'; Susan Hurley, *Natural Reasons* (Oxford University Press, Oxford, 1989).
13 Jon Elster, *Sour Grapes* (Cambridge, Cambridge University Press. 1983), p. 22.
14 Rawls, *Theory of Justice*, p. 143.
15 See Mellor, 'Objective Decision-Making'.
16 David Hume, *A Treatise of Human Nature*, ed. P. Nidditch (Clarendon Press, Oxford, 1968), p. 416.
17 Pettit, 'Decision Theory and Folk Psychology'.
18 Notice that this argument turns on the assumption that Parfit's condition of reflection, as mentioned in section III, is reductively specifiable.
19 See Gary Watson, 'Free Agency', *Journal of Philosophy*, 72 (1975), pp. 205–20.
20 Parfit introduces this second of two arguments in the course of providing the S-theorist with a reply to the first. The first argument is an implausibility charge of a kind that is familiar and, by his own admission, unprovable (pp. 130–6). It counters the S-theorist's claim that it is supremely rational to desire that things go as well as possible for oneself over one's life with cases where it seems rational for an agent to satisfy desires – say, moral concerns – that require the frustration of self-interest. This would tell against S in either background or foreground role. The S-theorist replies to this argument that he does not have to assume all three of the elements involved in S: viz. that the agent should promote (1) the good, (2) for himself, (3) over his life. He assumes the first two elements – that the good for the agent provides a reason – and he offers an argument for the third – that if the good for the agent provides a reason, then it does so in a time-neutral way. Parfit uses against this reply a principle that is later strengthened into the Appeal to Full Relativity. That appeal is meant to undermine the reply as well as to provide a second argument against S.

21 Thomas Nagel, *The Possibility of Altruism* (Oxford University Press, Oxford, 1970). Parfit distinguishes his own destabilizing argument from Nagel's (p. 144).
22 We are grateful for the useful discussion this paper received when it was presented at the annual conference of the Australasian Association of Philosophy, Sydney, 1990. We are also grateful for comments received from Frank Jackson, Lloyd Humberstone and Graham Oddie. We are greatly indebted to Jonathan Dancy and Derek Parfit for detailed comments on an earlier draft, and are also in Dancy's debt for helpful exchanges on the interpretation of *Reasons and Persons*.