In defence of self-interest: A response to Parfit

Simon Beck
Philosophy Department, Rhodes University, Grahamstown 6140, Republic of South Africa

Received March 1987

Parfit argues in *Reasons and persons* that acting according to your present desires is more rational, or at least as rational, as acting in your long- term self-interest. To do this, he puts forward a case supporting a 'critical present-aim theory' of rationality against the self-interest theory, and then argues against a number of possible replies. This article is a response to these arguments, concluding that Parfit's favouring of the present-aim theory is unfounded, and that self-interest is indeed the better theory of rationality.

Parfit betoog in *Reasons and persons* dat handeling in terme van huidige begeertes meer, of ten minste ewe, rasioneel is as handeling in terme van eie-belang op lang termyn. Ten einde dit te doen, ontwikkel hy 'n kritiese onmiddellike-doel-teorie' van rasionaliteit, en verweer homself dan teen 'n aantal moontlike besware. Hierdie artikel is 'n repliek op Parfit se argumente. Die gevolgtrekking is dat Parfit se voorkeur vir die onmiddellike-doel-teorie ongegrond is, en dat eiebelang die beste teorie van rasionaliteit bied.

Section 1

At the end of Part 2 of *Reasons and persons*[1], Derek Parfit offers us a choice of two conclusions: either that we reject S, the self-interest theory of rationality, or that we accept at least that S is no better a theory of rationality than CP, the 'Critical Present-Aim' theory. In the contest between these two theories, Parfit holds that the onus is on S to demonstrate its superiority over CP, in the light of a prima facie case in favour of the latter. He then presents a barrage of argument against the possible responses that S might offer, in an effort to leave S with no escape route.

Because of the large number of its component arguments, and the resulting complexity of the case against S, an attempt to defend this view can be a long and difficult task. However, I will attempt the task, examining many of the most important moves Parfit makes, and trying to answer the criticism he offers. I argue that in the end he has not provided adequate support for either of his two conclusions, and that S is to be preferred to CP.

S, the self-interest theory of what constitutes rational action, holds that one should do whatever is most in one's own interest, regardless of the effect this might have on others. Parfit acknowledges that there are a number of different views on what constitutes a person's best interests; for the most part of this article a 'desire fulfilment' theory will be adopted. That is, what would be in someone's best interests is what, throughout his life, would best fulfil his desires. (Throughout the discussion it is assumed that one does not desire to act in the interests of others rather than in one's own interest.) Thus S can be characterized as follows:

> S: It is rational for X to do whatever would best fulfil, or enable him to fulfil, his desires over his whole life.

CP, the critical present-aim theory, tells one that it is most rational to do whatever would best fulfil one's present desires. It is a *critical* present-aim theory because it excludes certain desires as intrinsically irrational, and may include others as rationally required — i.e. we would have reason to fulfil them even if we did not at present have them.

> CP: It is rational for X to do whatever would best fulfil his present desires, as long as these are not intrinsically irrational and do not conflict with any rationally required desire.

CP and S will conflict when the bias in your own favour is not one of your most pressing present desires, and the present desires you do have recommend action that is not in your long-term self-interest.

In choosing between CP and S as theories of rationality these conflict cases will be important. Parfit points out, however, that if the bias in one's own favour was rationally required in the sense outlined, then the two theories would never conflict. But to show that a bias in one's own favour is rationally required, he holds that the S-theorist must establish it to be the supremely rational desire — i.e. more rational than any desire which might go against it. This is the task Parfit sets for the S-theorist if he (the S-theorist) wants to retain his theory and not adopt CP instead.

Section 2

One great difficulty here for S, according to Parfit, is that there are other desires which are no less rational than S's bias, yet which conflict with it. For example, he suggests that it is just as rational to want to act in the interests of others when morally required to do so as to want to act in your own interests. This creates no special problem for CP since the latter theory can claim that the desire to act morally is rationally required, and one must thus act in the interests of others when morally required to, whether one wants to or not. But the possibility of there being a desire no less rational than the bias in one's own interest creates a fatal problem for S if S has to establish its bias as the supremely rational desire. If S has no acceptable way of showing the desire to act in one's own interests to be more rational than the desire to act morally, then CP is the theory of rationality we should adopt.

The desire to act in the interests of others is not the

only desire which Parfit believes to be as rational as the bias in one's own favour. It would be equally rational for an artist to want to paint as well as possible, or for a scientist to want to make some important discovery, even when these desires conflict with the self-interest of those people. Once again, CP makes room for the rationality of these desires by allowing that they may be rationally required (for these people), while they pose a direct threat to the truth of S. Parfit believes that S has no satisfactory answer to this criticism.

Before going on to look at Parfit's responses to other arguments in favour of S, this first criticism demands some attention. The question must arise as to just what Parfit is claiming for the various desires put forward, in saying that they are no less rational than self-interest. Parfit is not just testing these desires against our intuitions, claiming them to be intuitively just as rational. For he writes in section 50 that some of his claims about what is or is not rational will be intuitively implausible, but that is just because we are steeped in the tradition of the self-interest theory. Since our intuition is not to be trusted in these matters, Parfit cannot be appealing to any intuitive notion of rationality with regard to which these various desires are equally rational.

He can also not be appealing to any substantive notion of rationality, for that is what S and CP are attempting to provide.It would be only after we had decided to accept one such theory of rationality that he could make any such claim.

The only plausible alternative seems to be some 'minimal' or 'thin' theory of rationality. Such a theory would be that which is common to and underlies all substantive ('thick') theories, in the way in which John Rawls[2] envisages a thin theory of the good underlying the relevant thick theories.[3]

Quite what the thin notion of rationality would be, Parfit does not make clear (again, our intuitive notions are not to be trusted!). But he does talk of a theory of rationality aiming at telling us what we have most reason to do, and he rejects as irrational certain desires which 'could not provide us with a reason for acting'(p.120). Desires must at least be able to provide reason for acting if they are to be candidates for inclusion as rational desires. This is, while not much, at least the start of a thin theory. In the light of this thin theory, a rational desire would be one which provides us with a reason for acting. This would have to be the notion operating in Parfit's claim that the proposed desires are no less rational than the bias in one's own favour.

Before discussing whether some thin theory of rationality can perform the task assigned to it here, something needs to be said about Parfit's notion of a reason for acting. This is a notion which he does not make sufficiently clear, despite it being an extremely important one. He is quoted above as saying that certain desires *cannot* 'provide us with a reason for acting'. On first sight, this appears counter-intuitive. For if reasons for actions are understood (following Davidson)[4] as the beliefs and desires which cause those actions, then there seems to be nothing in the way of any desire 'providing us with a reason for acting' (no matter how outlandish it may be).

Parfit's use of the term 'reason' needs to be handled very carefully as a result of this. He is not using it in the sense suggested. Rather, when he talks of a reason for acting, he is talking of the beliefs and desires (or whatever it is) which would make that act rational. 'Rational' in this context is an evaluative term. An action is not rational simply because it is caused by the beliefs and desires which 'rationalize' it (in the Davidsonian sense); it can be so caused and yet be irrational. It is rational if it meets with certain evaluative criteria — Parfit talks of rational acts as one talks of moral acts: he is concerned with justification rather than causation.

So a desire which could not provide us with a reason for acting is one which could not possibly justify an action (even if it did cause that action). And when Parfit talks of X's reason for doing A, he means X's belief/desire which one would use (or which X would use) to justify that action — not necessarily the belief/desire which caused X to do A.

To return to the thin theory of rationality: although appeal to a thin theory is the only plausible way to fill out the notion of rationality Parfit is making use of, even it will not fill the role Parfit requires it to play. For Parfit demands that the S-theorist show his bias to be the supremely rational desire. Using only the thin theory of rationality, however, this is an impossible demand. It is impossible because the thin theory is, by its very nature, neutral between competing substantive theories of rationality; it consists of what is common to all substantive theories, and thus cannot decide between them. We can only use it to judge whether desires are rational in a weak sense — whether they provide a reason for acting — we cannot use it to decide what the most rational desire is. Only a thick theory can be used for that.

If this is correct, Parfit's demand that the S-theorist must show his bias to be the supremely rational desire would be out of order. S provides a substantive definition of 'rational' in terms of this very bias, and it is thus logically impossible that there be a conflicting, but equally rational, bias. The bias in one's own favour would be in a strong sense 'rationally required' (although only after S is accepted as the thick theory of rationality).

This line of argument raises a question about the adequacy of Parfit's CP as a theory of rationality, as well as showing his demand on S to be an impossible one. For it is using only the thin notion of rationality that the concepts of 'intrinsic irrationality' and that of a desire's being 'rationally required' must be explained; I shall argue that no acceptable thin theory is capable of explaining these notions as Parfit understands them and uses them in his CP.

A thin theory of rationality is quite capable of offering a weak notion of intrinsic irrationality, and thus ruling certain desires out of contention. For instance, it would presumably class as intrinsically irrational self-contradictory desires such as a desire not to fulfil any of one's desires. We could also exclude here desires which cannot be satisfied and, perhaps, desires for which no reason can be given or which have no point. All these desires appear to be desires which no thick theory of

rationality will tolerate, and thus their exclusion from the class of rational desires will make up part of the thin theory. The problem is that Parfit's notion of Intrinsic Irrationality is not the weak one outlined here.

An intrinsically irrational desire, according to Parfit, is one whose object is 'in no respect worth desiring, or is worth avoiding'(p.123). He gives the example of one who desires to feel great pain (although not for masochistic 'or any particular reasons') (Note: It is true that Parfit could rule this desire as intrinsically irrational using the thin theory. He could say, for example, that it has no point. However, it is not the example but the principle it is intended to illustrate that is at stake here.) Parfit's CP, then, presupposes some theory which spells out the notion of what is 'worth avoiding'; his notion of intrinsic irrationality is not a weak one.

Now, if 'worth avoiding' is to be spelt out in terms of what is not in one's self-interest (which seems to be the intuitively obvious way of understanding it), then CP collapses into S. The alternative is that we spell it out in terms of some moral theory, or some other theory of rationality. But then we have two thick theories of rationality, CP and the theory it presupposes (If a moral theory was chosen to perform the task it would become a theory of rationality, since it explains what makes certain desires irrational). The presupposed theory is presumably the more fundamental one, and yet it is a very odd one. If 'worth avoiding' is explicated using some moral theory, as it seems must be the case, then this presupposed theory will hold that it is always irrational to want something bad. If this were the case, then S would obviously be wrong. But this is nobody's view of rationality, and is quite implausible. One of the very reasons for talking about theories of rationality as opposed to limiting discussion of motivational theories to moral ones is that it appears obviously true that we often have reason to want and do things against the tenets of morality. These considerations demonstrate the internal deficiencies of Parfit's CP,leaving us unable to give an adequate account of notions crucial to it. Because of these deficiencies, CP loses out as a rival to S.

Section 3

Parfit acknowledges that S has other resources to support its claim of superiority over CP, should the S-theorist acknowledge the force of Parfit's 'first argument'. He foresees the S-theorist making the claim that reasons are *temporally neutral,* arguing that 'the force of a reason extends over time'.That is, the S-theorist will argue that since you will have reasons in the future to fulfil your future desires, you have such reasons now. 'What you have most reason to do, is whatever would best fulfill, or enable you to fulfill, all of your desires throughout your life' (p.137). The truth of these claims would entail that presently held desires are not the only rationally significant ones, and thus that CP is a fundamentally misguided theory.

But this argument leads to problems: if the S-theorist held that the force of a reason extends over time, he would be bound to act according to certain past desires which he no longer has, and, according to Parfit, this leads him into absurdity. Also, in the case of desires one knows one will have in the future but which are based on beliefs one now thinks false or contemptible, one will be bound to giving equal weight to these desires as one gives to present desires based on beliefs one thinks true. Again, Parfit suggests that this is an absurd position. Furthermore, he thinks that arguments S puts forward against reasons being time-relative can be used to argue against reasons even being 'agent-relative'. Thus, in cases where S and moral theory conflict, S would lose out to moral theory which makes this latter claim.

The first example of the absurd consequences of following S emerge in Parfit's 'Saving Venice' example (p. 152). For 50 years someone has contributed to the Save Venice fund in order to fulfil his two strongest desires (i) that Venice be saved, and (ii) that he be amongst its saviours. Now he loses these desires, or they are overridden. His most pressing present desires are to put the money he would have contributed into other channels — into cultivating a rose-garden, let us say. Would it be more rational for him to contribute to the fund or to grow roses? The S-theorist must work out what would best fulfil the person's desires over his whole life; assuming that the fulfilment of strong desires is more important if they are long lasting, and given that the force of a reason extends over time, the S-theorist must say that the rational action is to contribute. And this, says Parfit, is absurd.

There are two points in the above argument where Parfit's case is extremely weak. The first is that the absurdity of the S-theorist's position only emerges once it is assumed that long-standing desires are to be granted more weight than more ephemeral ones. For it is only on this assumption that the desire to contribute to the Save Venice fund will take precedence over the S-theorist's more pressing current desires, and thus that fulfilment of the former desire will become the more rational course of action. Now, it is by no means clear that one should have more reason to act on a desire simply because one has had it for a long time — it is only with regard to a certain kind of desire, such as the desire to stop smoking, that the longer one has it, the more reason one has to fulfil it. Parfit's argument is then perhaps better seen as a *reductio ad absurdum* of the assumption that the fulfilment of long-lasting desires is especially important than as an argument against S.

The second point is the more important one. It concerns the question of exactly what the Venice-saver's reasons for contributing are (bearing in mind the discussion of 'reason', *supra*). As Parfit outlines the example they are the desires that Venice be saved and that he (the Venice-saver) be among its saviours. These are the reasons whose force Parfit insists the S-theorist takes as extending over time and that is where the difficulties arise. But where Parfit goes wrong is that they are *not* the Venice-saver's reasons. His reason for contributing, while he still has the desire to contribute, is that it is in his interests (i.e. it fulfils his desires) to contribute. A reason for an action for Parfit in this context, it must be remembered, is whatever it is that makes the act

rational. And it is this (the action being in one's interest) which makes any action rational according to S. If one insists on characterizing reasons as beliefs or desires, then it is the belief that doing A is in one's interests, or the bias in one's own favour, which is one's reason for doing A.

Once this is realized, Parfit's case falls apart. For, following the desire-fulfilment version of S, the reason for X performing any action will be that it would contribute towards X best fulfilling his desires *over his whole life.* As a result, the force of the Venice-saver's reason for contributing *necessarily* extends over time (as, indeed, does that of the reason for any action, according to S). It also becomes clear that it would not be rational for the Venice-saver to continue donating money once he loses the relevant desires; since it would no longer go towards best satisfying his desires over his whole life, he would no longer have any reason to do so.

Parfit's second objection to the temporal neutrality of reasons was that it may require one to give the same weight to future desires based on beliefs one now thinks false as one gives to present desires based on beliefs one thinks true. To do so would be absurd, he says. But his support for this claim is vague; why should it be absurd to give equal attention to such desires?

Take the example of a radical student who is advised by S not to take part in any activity which might damage his future career with IBM, and which he will regret in years to come in his management position with that firm. In other words, according to S he has reason to take his future desires based on ideals he now finds contemptible into account. CP would tell this student to refuse the job with IBM, and indulge in whatever activity fits in with his ideals at the moment. Now, without opting for one of the competing theories on independent grounds, we cannot say which is the more rational of the two courses of action; but there is nothing apparent which shows the former course to be absurd. All that the student is asked to do by S is to take a realistic attitude to the way in which he is likely to change, and there's nothing absurd about that.

The final objection was that an analogous argument to that used by the S-theorist against temporal relativity could be used against S itself. For if reasons need not be relative to a time, why need they be relative to a person? S says that you have reason to act in your interests regardless of the effect this may have on others, and others have reason to act in their interests regardless of what the consequences of such action may be for you. But on Parfit's view of moral theory, on the other hand, what is a reason for one is a reason for all: it is 'agent-neutral', not assigning different reasons for acting to particular individuals (for example, if one person has reason to relieve someone's suffering, then eveyone has reason to relieve that person's suffering). Parfit's suggestion is that if one rejects one sort of relativity, one will be bound to reject all sorts. S, in rejecting the relativity of reasons to time, would have to reject the relativity of reasons to persons as well.

He supports the suggestion by arguing that 'I' and 'now' are formally analogous, and ought to be given the same treatment. Thus if the S-theorist gives himself special treatment, then he should also accord special treatment to the present. 'If reasons can be relative, they can be fully relative: they can be relative to the agent at the time of acting,' he says (p.140). Conversely, if S refuses to accept relativity to time, he should refuse to accept relativity to persons as well. As a result, S is trapped between CP and moral theory, with both of which it conflicts: by rejecting the tenets of one by appeal to, or argument against, relativity, it plays into the hands of the other.

This argument will not work, however. It rests upon the formal similarity of 'I' and 'now'. But the fact that these two words are indexicals, and thus formally similar, does not provide any grounds for demanding that persons and times be treated in the same way when it comes to the scope of reasons. Parfit needs a strong argument as to why partial relativity of reasons is unacceptable and must give way to full relativity, but this is not forthcoming.

Section 4

Having supposedly rid himself of the S-theorist's 'second argument', Parfit turns his attention back to the first argument, and piles on more objections against any attempt by S to establish the bias in one's own favour as the supremely rational desire. One of the desires S must defend its bias against is the bias 'towards the near', he says.

> *The bias towards the near* is the preference for nearer pleasures simply because they are nearer (and for further pains over nearer ones, simply because they are further).

This is a bias obviously closely connected with CP. Is the bias in your own favour more rational than this? As I have argued in Section 2, a thin theory of rationality would not be adequate for providing an answer to this question; but Parfit believes that S might try an argument independent of thin theories to establish its superiority.

Parfit expects the S-theorist to argue against the rationality of putting off pains until later even at the risk of making them worse, a course of action advocated by the bias towards the near. The best method available to S for this purpose, according to Parfit, is to argue that a pain hurts just as much whether you experience it now or put it off until later. But he holds that once again the S-theorist's argument will render him vulnerable to attack by moral theory. In this case, the moral theorist can use a closely analagous argument to S's own — he can say that a pain hurts just as much whether it happens to you or to someone else; and this would be an effective argument against the bias in one's own favour. In this way, S's best response to the bias towards the near backfires.

Parfit will not allow S to reply, 'yes, but the reason why Proximus (the man who holds the near-bias) is incorrect is that a pain hurts *you* just as much regardless of when it happens; whether a pain is further in your future or not it is no less painful to you, but if a pain is

someone else's, then it is less painful to you.' If the S-theorist wants to argue this, Parfit contends that he must first establish that it can be of rational significance *who* feels a pain.

But Parfit's case collapses here. S is called upon to criticize Proximus and his CP cronies if he is to defend S against CP — and Proximus, in so far as he accepts CP, already accepts the rational significance of who feels a pain. S is thus entitled to appeal to it, and, as spelt out above, it is a crucial part of the argument against Proximus: for there is nothing irrational in preferring someone else's having a pain later to your having one now. So Parfit cannot respond by arguing that the insertion of 'you' into the claim that a pain hurts just as much whether it occurs now or later is redundant. And when S includes this extra word, morality loses the ability to use an analagous argument against it: it is nonsense to claim that a pain hurts you just as much no matter to whom it happens! The only onus of proof in the argument between Proximus and S is not on S, but on Proximus and other associates of CP, and that is to establish the rational significance of time in such cases as the one above. The argument Parfit has set up against S fails.

Section 5

In the event of S's argument against Proximus not working, as Parfit (mistakenly) believes it does not, he thinks that S might once more attempt to appeal to temporal neutrality to overcome the bias towards the near. He holds that the appeal will fail once again; but I will argue that even though the S-theorist does not need this way out, it is available to him.

The earlier argument against temporal neutrality (Section 3) was aimed at 'desire-fulfilment' versions of S, but other versions remain unscathed. The alternative versions Parfit has in mind would spell out one's self-interest as what would make one's life happiest (Hedonistic theories), or in terms of the things that are good or bad for us ('Objective list theories').

Assuming we opt for one of these versions, if the force of a reason extends over time, then Proximus would be irrational to follow his bias when it conflicted with his self-interest. Parfit has new complaints against temporal neutrality, however. The problem concerns what Parfit calls the bias towards the future:

> *The bias towards the future* is the preference of having undergone a great pain in the past to having to undergo a lesser pain in the future.

Temporal neutrality demands the irrationality of this bias, but in Parfit's eyes it is by no means clear that it is irrational. Thus the appeal to temporal neutrality in order to defeat Proximus could ultimately be an embarrassment to S.

If the S-theorist is to retain temporal neutrality, then he will have to find some way of justifying the bias towards the future and which excuses it from meeting the requirements of such neutrality. Parfit suggests that S may appeal to the passage of time to achieve this. S can argue that since time passes our past suffering need not matter to us, while our future suffering should: our relation to the future is different from our relation to the past. Because of this it would be irrational not to be relieved that pains are in the past. In this way, S can apparently retain the bias towards the future while rejecting the bias towards the near.

Why this argument is not available to S, Parfit argues, is that it has the following unfortunate consequences for him. Suppose someone hears that his fatally ill mother will suffer great pain before she dies. Naturally, he is distressed. Then he learns that this was partly incorrect — she is already dead, but did suffer great pain. This person would not be irrational in still being distressed (says Parfit). If he was to learn that his mother survived but had gone through great pain, he would still be distressed (though less so), and not without reason. But the appeal to the passing of time was to establish that concern at past pains is irrational; and so the S-theorist is bound to argue that this person's distress is irrational. In this way the S-theorist's defence of the future-bias leads to counter-intuitive results. As a result, S is left without an argument against the bias towards the near, and this counts against the bias in one's own favour being the supremely rational desire. Parfit suggests that there are other ways in which S may attempt to defend its bias, but argues that none of these will work (sections 71 & 73).

But this argument against S's appeal to time's passage is not convincing. Firstly, Parfit sees too much in the bias towards the future. As it was characterized above, it is fine, but Parfit infers from our preference for greater pains in the past over lesser ones in the future that S must hold it irrational to be concerned about past pains. (The jump from talk of *our* pains to talk of pains in general is too quick, but that is not my present point.) This, however, just does not follow — as characterized, the future-bias is a mere *preference*; it thus rules out a certain preference (namely, preferring future pains to past ones) as irrational, not concern at past pains. S is unlikely to argue that such concern is irrational in itself: we see nothing irrational in avenging the harms others have caused us (though we may see this as immoral), it may well be in our interests to take some revenge, motivated partly by concern at our past pains. And this need not conflict in any way with our future-bias. All the appeal to time's passage is required to do, then, is to support our preference towards the future — it need not go as far as Parfit makes it go, and support a lack of concern about the past. And by not ruling out all concern at past pains, the S-theorist does not lay himself open to Parfit's counter-example.

Section 6

None of the arguments examined above give us reason to prefer CP to S as a theory of rationality. One of Parfit's alternative conclusions was that we have a tie between the two theories. My distinction between thin and thick theories of rationality suggests that we can only expect a tie between competing thick theories when we do not appeal to intuition or the internal deficiencies of a

thick theory. I have argued, however, that CP does suffer from such deficiencies. As a result, there is no tie between CP and S; rather S is the theory to be preferred of the two.

Acknowledgement

I am grateful to the HSRC for financial support.

Notes

1. Oxford, Clarendon Press, 1984. All page references in the text are to this book.
2. *A theory of justice.* Oxford: Oxford U.P., 1972, Section 60.
3. In *Sour grapes* (Cambridge: Cambridge U.P., 1983) Jon Elster makes use of a thin theory of rationality, also appealing to the notion introduced by Rawls. As will become clear, however, what I am calling the thin theory of rationality differs from Elster's version.
4. Davidson, D. Actions, reasons and causes, in D. Davidson, *Essays on actions and events.* Oxford: Clarendon Pr., 1980.