The Present-Aim Theory of Rationality*

Shelly Kagan

In the second part of *Reasons and Persons*,[1] Derek Parfit offers an amazing array of arguments concerning rationality and time. The discussion is sophisticated and intricate, touching on a variety of topics. The central concern, however, is with the defense of one particular theory of rationality, the Present-aim theory. In this paper I want to examine the adequacy of the three major lines of argument that Parfit offers to establish the superiority of that theory over a rival view, the Self-interest theory. Not surprisingly, however, it will first take a bit of stage setting to get us to the point where we can evaluate these arguments. Bear in mind, in what follows, that my way of setting out the problem does not correspond exactly to Parfit's.

THE ISSUE

Imagine that I am presented with a full matrix of information, listing all the desires of all people, past, present, and future. Each desire is indexed to indicate *whose* desire it is (whether mine, yours, or Fritz's), and *when* it occurs (whether now, yesterday afternoon, or twenty years from now). Given this wealth of information, what is it rational for me to do? Which desires should be given weight in my deliberations? And how much weight?

Different theories of rationality offer different answers to this question. Theories differ, for example, as to whether I should give direct weight only to my own desires, or as to whether my future desires should count as heavily as my current ones. Corresponding to each theory of rationality, in effect, there is a function, assigning weights ranging from one to zero to each of the desires in the matrix. Approximately following Parfit, we can distinguish three prominent theories.

The Self-interest theory, S, gives full weight to each of my own

* I owe an enormous debt to Derek Parfit for many illuminating discussions of the issues examined in this paper.

1. Derek Parfit, *Reasons and Persons* (New York: Oxford University Press, 1984). This paper was originally presented on March 22, 1985, at the Pacific Division meeting of the American Philosophical Association, in a symposium on *Reasons and Persons*. All page references in the text are to this work.

Ethics 96 (July 1986): 746–759

desires (regardless of when it is held), and no direct weight to the desires held by others. I say "no direct weight" because if one of my own desires is to see to it that, for example, my mother is happy, then her desires will come into my deliberations indirectly; but they will not have any weight in their own right. According to S, were it not for my concern for her, my mother's desires would not enter into my deliberations. Each of my own desires, however, directly generates a reason for acting, and the most rational act overall depends on the balance of the reasons generated by my desires. Note that the strength of the reason generated may depend on such factors as the intensity of the desire, but it does *not* depend on when the desire is held: my future desires, for example, are given just as much weight in their own right as are my present desires.[2]

In contrast to S, Morality gives direct weight to the desires of others. Different moral theories disagree about the details, of course, but for simplicity let us stick to a version of Utilitarianism which gives equal weight to all desires in the matrix, regardless of whose desires they are, or when they are held. According to this Utilitarian version of morality, U, each desire generates a reason for my acting on it—even though the desire in question may not be my own and I may not care about the person whose desire it is.

Finally, we have the Present-aim theory, P, which is like S in giving weight only to those desires that are held by me. Unlike S, however, it adds the further restriction of giving direct weight only to those desires that are held by me *now*. On this theory, only my current desires directly generate reasons for my acting. The desires of others, or my own future desires, are assigned no weight in their own right; they can only indirectly generate reasons, and even that only if I am currently concerned about my future or the welfare of others.

There are, of course, many other theories of rationality. I have not mentioned, for example, any theories that give only *partial* direct weight to certain desires. But we can limit our discussion to these three. There are other complications I want to put aside as well: each of S, P, and U can come not only in "instrumental" versions of the sort I have described but also in "deliberative" versions that assign weight only to desires that survive some process of reflection, or in "critical" versions that rule out certain desires as intrinsically irrational or demand other desires as rationally required.[3] Furthermore, it is only on certain conceptions of self-interest

2. For simplicity, I will hereafter disregard the variations in weight that are due to the intensity of the given desire. Furthermore, in order to avoid the complications that arise from the possibility of misinformation on the part of the agent, I will assume that we are considering "objective" theories of rationality—i.e., theories about what the agent has reason to do, whether or not he is in a position to realize it.

3. When Parfit makes these distinctions (p. 94, and secs. 45–46), he only brings them out in terms of P. He fails to note that other theories of rationality can come in these different versions as well. This sometimes leads him to underestimate the potential strength of one of these rival theories. See, e.g., n. 11 below.

that what I have described as S can properly be called a "self-interest theory" at all.[4] All of these issues merit attention, but in the interests of time I am going to leave them aside.

It is worth emphasizing that each theory of rationality needs a "master" function telling the agent how much weight, if any, to give to the various desires represented in the matrix. Theories will disagree about what function is correct, but they cannot avoid commitment to some function or the other. A given theory will view the agent as acting rationally when he performs the act supported by the balance of reasons—but it is the master function corresponding to that theory which determines which desires generate reasons in the first place.

It is also worth stressing that S, P, and U are in themselves neutral as to what particular desires the agent may have or lack (although critical versions of the theories may not be). As I have noted, for example, nothing in S rules out my having a desire that my mother be happy: this desire gets the same weight as any of my other desires. The master function of a theory of rationality tells me how much weight to give to desires—including my own—but so long as my actions conform to the weightings endorsed by the function, the theory itself does not care what desires I may have or lack.

Of course, if I do conform to the master function of a given theory, particularly if I do this knowingly, it will no doubt typically be appropriate to ascribe to me a *desire* to so conform. We might call this a *metadesire,* for it is a desire concerning the direct weight to be given to the various ordinary *n-order* desires. Thus, someone who conforms to S can typically be said to have a metadesire to give direct weight only to his own desires. Similarly, someone who conforms to P can typically be said to have a metadesire to give direct weight only to his own current desires. And so on. Ascriptions of such metadesires are not always unproblematic, but, once again, I will pass over these complications.[5]

4. On certain conceptions of "self-interest," some of my rationally acceptable desires may not be relevant to my self-interest, for satisfying them will have no necessary connection to making my life go better (see app. I). On such accounts, what I have described as S is not a genuine self-interest theory, for it gives direct weight to *all* of my desires (whenever they are held)—and not only those relevant to my self-interest. However, regardless of whether this version of S deserves the title "self-interest theory," it is an important rival to the Present-aim theory, which gives direct weight only to my *current* desires.

5. There are cases in which possession of the desire to conform to a given theory of rationality will lead the agent to fulfill that theory less well than if he lacked the desire in question. (See Parfit's discussion, in chap. 1, of theories that are "indirectly self-defeating.") Suppose that, in keeping with such a theory of rationality, the agent *lacks* the conscious desire to conform and so conforms more fully. Ascription of the metadesire in such a case seems problematic. Even in those cases where the metadesire can be ascribed, there is the question of whether it falls under its own scope. Does the master function treat the metadesire as one more of the agent's current desires, to be weighted accordingly? Typically, there is no problem if it does. But for what we might call "perverse" master functions—ones which instruct agents to frustrate their own desires—taking the metadesire to fall under its own

The ascribability of the appropriate metadesire can lead us into some tempting confusions. For example, it is sometimes thought that someone who conforms to S must be selfish, not caring for others. And as we have just seen, we can in fact ascribe to such a person the metadesire to give direct weight only to their own desires. But as we have also seen, this is completely compatible with the person having numerous *n*-order desires for the well-being of others. The master function of S will insist that these desires be given as much weight as any other desires the agent possesses, and doing so will satisfy the metadesire as well. Thus one who conforms to S need not be selfish.[6]

The presence of the metadesire may also tempt us into thinking that conformity to the given master function is rational *by virtue* of the fact that it satisfies the metadesire. In fact the reverse is closer to the truth. For knowing that an agent has a particular metadesire will not tell us what he has reason to do unless we make an independent assumption about which desires (including metadesires) generate reasons. And to do this is simply to assume the soundness of some particular master function. To put the point another way: desires do not generate reasons by virtue of their satisfying the metadesire; at best this provides an extra indirect boost. Rather, desires generate reasons by virtue of the master function. A theory of rationality cannot avoid commitment to some particular function.

As I have noted, the functions of S and U differ on the issue of whether the desires of all persons, or only the agent's own desires, generate reasons for the given agent. Theory S gives direct weight only to the agent's own desires; it is agent-relative. Theory U, in contrast, is agent-neutral: it gives direct weight to the desires of all persons. Thus, U accepts, and S rejects, personal neutrality.

Similarly, it seems clear that the issue between S and P is the acceptance or rejection of temporal neutrality. The two theories agree that reasons are generated only by desires of the agent, but P holds that only the agent's *present* desires do this, while S holds that *all* of the agent's desires do this, whether or not they are currently held.

Given this, it is somewhat surprising to find Parfit insisting that temporal neutrality is *not* the issue between S and P, and he devotes a section (52) to arguing this point. However, I *think* that the appearance of disagreement here is misleading. I take Parfit's point to be that there is nothing in P which rules out the possibility that an agent's desire concerning her own life be a temporally neutral one: that is, insofar as she cares about her life, she is equally concerned about all portions of

---

scope can lead to paradox (similar to the paradox of the liar). I must leave these fascinating issues aside.

6. The ascribability of the metadesire corresponding to P also explains the temptation to think that P must be committed to what Parfit calls "Egoism of the Present" (see pp. 134–35; cf. p. 142).

it. Indeed, there is a critical version of P in which the temporal neutrality of this desire is demanded (pp. 135–36). But this only shows that P can accommodate temporal neutrality as far as the *content* of the *n*-order desires is concerned. None of this affects the point that for all versions of P it is only *present* desires which directly generate reasons. Theory S, in contrast, holds that *all* of an agent's desires generate reasons, regardless of when they occur. For S, time is irrelevant; for P, it is paramount. In this sense, temporal neutrality is indeed the issue between S and P.

Ideally, a defense of P over S would justify the sensitivity P displays to the question of when the desire is held. We would like an explanation of why such temporal relativity is appropriate to a theory of rationality. Unfortunately, Parfit does not provide such an account. He does, however, offer several arguments in defense of the claim that P is indeed the correct theory of rationality, as opposed to S. Showing this would certainly be a significant accomplishment. Therefore, with our stage setting in place, we can at last turn to the evaluation of these arguments.

PARFIT'S FIRST ARGUMENT

In its essentials, Parfit's "first argument" needs only two steps. Let us introduce the expression "the bias in one's own favor" to refer to the desire to give equal weight to all of one's own desires and no direct weight to the desires of others. The first step of Parfit's argument is the claim that the S-theorist must hold that this bias is supremely rational—that we should not care as much about anything else. The second step of the argument, then, is the suggestion that it is not, in fact, plausible to hold that no other desires are as rational as the bias in one's own favor. On the contrary, numerous "patterns of concern" are just as rational as the bias, for example, a concern for the interests of others, a commitment to being moral, various desires for achievement, and so on. Theory S, therefore, needs to be rejected (secs. 50, 51; cf. pp. 146, 147).

In effect, Parfit's first argument comes to this: S is necessarily intolerant in its attitude toward the rationality of different patterns of concern. It implausibly elevates one particular pattern—the bias in one's own favor—and gives it a unique theoretical status. The argument against S is thus fairly direct. On the other hand, the support this argument provides for the Present-aim theory is relatively indirect. It is clear, however, that Parfit believes that P can be tolerant where S must be intolerant: P can accept the rationality of numerous patterns of concern and instruct the agent to act in conformity with whatever pattern happens to reflect his strongest desires. If we are drawn to the pluralistic attitude toward patterns of concern, we have reason to reject S and to accept P.

Whatever the attractions of Parfit's first argument, I believe it should be rejected. For the most part, the pluralistic intuitions appealed to in support of the second step of the argument are not ones that need to trouble the S-theorist; and to the extent that S can be convicted of intolerance here, P stands similarly convicted.

Theory S, of course, requires the agent to act in conformity with its own particular master function—giving direct weight only to his own desires. And, as we have noted, an agent who conforms with S can typically be said to have the corresponding metadesire to weight desires in this fashion. If we wish, we can say that S requires that an agent have the bias in his own favor as his metadesire. No *other* metadesire will be appropriate, for no other metadesire will accurately portray the reasons generated by different desires. To this extent, the first step of Parfit's argument is correct: S does elevate one particular desire to a unique theoretical status.

However, what Parfit seems to have overlooked is that P also unavoidably elevates one particular desire to a unique theoretical status. (Indeed, so does almost every theory of rationality.)[7] For P requires that the agent act in conformity with its *own* particular master function—one in which an agent gives direct weight only to his own current desires. An agent who conforms to P will have as his metadesire the bias in favor of his current desires. And the P-theorist, in turn, must view this metadesire as uniquely appropriate.

None of this, however, runs afoul of the intuitions that various desires for achievement, or concern for others, and so on, are perfectly rational. As I have stressed, neither P nor S needs to rule out any particular n-order desire as rationally unacceptable or inferior. If, for example, I have a desire to help others, then according to P this generates a reason. And if this desire is sufficiently strong, the reason it generates may override my other current n-order desires. But no matter how strong my concern for others may be, that desire cannot have the same theoretical role as my metadesire to give direct weight only to my own current desires. For were it to play the role of my metadesire, I would no longer be conforming to the master function of P: I would be giving *direct* weight to the interests of others, in violation of the Present-aim theory.

The situation is similar for S. If at some point in my life (whether currently or not) I will have a concern for others, this generates a reason now. And if this concern is sufficiently strong, the reason generated may override other n-order desires, including my current desires. It is true, of course, that no matter how strong this concern for others, the S-theorist will never be willing to give this concern the theoretical role played by my metadesire, the bias in my own favor. But this is just to notice that S, like any other theory of rationality, must dig in its heels and insist that there is a uniquely correct metadesire—the one that reflects the correct master function. It certainly does not show that S needs to be unacceptably intolerant of the rationality of various desires, so long

7. There might, I suppose, be genuinely tolerant theories, which consider conformity to any one of several master functions equally rationally acceptable. (If so, my earlier remark that all theories are committed to the existence of a uniquely correct master function would need to be qualified.) This tolerance, in turn, would be reflected on the level of metadesires. Note, however, that P is not such a theory.

as it is kept clear that what is being conceded is the rationality of these different desires in their appropriate role as n-order desires.

Parfit might retort that tolerance on this level is not sufficient, that it is reasonable to demand a pluralistic attitude toward metadesires themselves. But this does not seem plausible. To accept various metadesires as equally appropriate is to hold that there is no single correct master function—no single truth of the matter as to which desires directly generate reasons. This is a view that few theories of rationality (if any) can accept. As I have emphasized, it must be rejected not only by S but also by P.

Why is it that Parfit overlooks the parallel between S and P—that is, the fact that both are tolerant on the level of n-order desires and intolerant on the level of metadesires? I think it may simply be because Parfit *believes* P. He believes that one's current desires generate reasons directly. This being so, the presence of a metadesire to give weight to those desires seems superfluous—as indeed it is. But given his commitment to P, Parfit cannot see how future desires will generate reasons except through the presence of a current concern for one's future. Thus, future desires will not get the weight that S gives them, unless S unreasonably demands that a temporally neutral bias in one's own favor be one's strongest desire.

If I am right about this, it should be clear that the discussion begs the question against S at a fairly deep level. A similar question-begging argument could have been offered by one committed to S. Given the belief that all of one's desires generate reasons directly, the presence of a metadesire to give weight to future desires will seem superfluous—as indeed it would be. And to the S-theorist it will be unclear how P can guarantee that satisfying my current desires will be rational, unless P unreasonably demands that I have an extremely strong bias in favor of my current desires.

Both S and P insist that there is a uniquely appropriate metadesire. But neither theory believes that it is the presence of that metadesire that makes conformity to the given theory rational. And so neither theory need insist that the metadesire be one's strongest desire. Thus Parfit's first argument should be rejected.

THE APPEAL TO FULL RELATIVITY

Parfit's second argument against the Self-interest theory (secs. 55, 57, 58) turns on the realization that S is only partially relative. In this it is unlike both the Present-aim theory and the Utilitarian version of morality that we have considered. On the one hand, U demands both personal and temporal neutrality: it gives direct weight to all desires, regardless of whose desire it is, or when the desire occurs. In P, on the other hand, both personal and temporal neutrality are rejected: P gives direct weight only to the agent's own desires (unlike the desires of others), and indeed only to the agent's current desires (unlike his future desires). Thus both U and P are "pure": in U, reasons are relative to neither the time nor the agent, and in P, reasons are relative to both.

In contrast, S is a "hybrid." It rejects personal neutrality yet demands temporal neutrality, insisting that although only an agent's own desires generate reasons for him, *all* of his desires generate reasons and not only his current desires. Thus S is incompletely relative. The driving force behind Parfit's argument—the "appeal to full relativity"—is the thought that either relativity is acceptable in reasons or it is not. If it is unacceptable, then both S and P need to be rejected, and only U survives. But if relativity is acceptable, then there is no justification for stopping at only partial relativity, so we must reject S and move to P.

But what, exactly, is it that is supposed to be problematic about incomplete relativity? At one point Parfit claims in passing that S "can be charged with a kind of inconsistency" (p. 140). If he means that there is something logically inconsistent about endorsing agent relativity while rejecting temporal relativity, then Parfit is clearly wrong. If, more charitably, he simply means that S is not constant in its attitude toward relativity—embracing one kind but not another—then this is correct, but why is it objectionable? Why must we move to full relativity once we have admitted the desirability of any kind of relativity at all?

The only thing I can find in defense of this view is Parfit's discussion of a certain formal analogy between oneself and the present (p. 140; cf. pp. 142–43). When I am deciding how to act, I am deciding what *I* should do. But, obviously, I am also deciding what I should do *now*. Given this analogy, one might conclude that reasons should be relative not only to the agent but also to the time of acting.

The formal analogy certainly holds, but it is hard to see how it helps the P-theorist make a case for the substantive thesis that reasons should be doubly relative. It is true that my act or decision occurs at a particular time. But it hardly follows trivially from this that my act should be based solely on what desires are *held* at that time. After all, it is also true that my act will affect what results at some later time—call it "then"—yet the P-theorist hardly wants to admit that my act should be based solely on what desires will be held *then*.

What is more, since the formal analogy between "now" and "then" is quite tight, if we were genuinely impressed by the fact that I act now to affect then, perhaps we should conclude that my act should be based on all of my desires that will be affected at all—desires held now or then. That is, we should accept S. Or going even further, since *I* act *now* to affect myself and *others* both now and *then*, if we were truly impressed by the analogy we might conclude that my act should be based on all the desires that it will affect, regardless of whose they are and when they occur. In short, we might accept U.

More plausibly still, we should reject the analogy as being inadequate to settle substantive disputes over the relativity of reasons. For the presence of the formal analogy is completely compatible with there being objective grounds for treating the two dimensions of person and time differently. In particular, it might be held that the fact that desires come bundled into unified lives is a rationally relevant consideration, justifying agent

relativity and temporal neutrality. Part 3 of *Reasons and Persons* is, of course, devoted to an attack on exactly this position, and I will not evaluate the merits of those arguments here. But Parfit believes that the appeal to full relativity can stand on its own, and this seems incorrect. Other than the inadequate appeal to the formal analogy, I do not see what grounds Parfit gives us for thinking it objectionable that S is only partially relative. Why must reasons be fully relative if they are relative at all?[8]

Although he is drawn to the position that reasons *must* be fully relative, Parfit recognizes that S is threatened even by the more modest claim that reasons *can* be fully relative. For S holds that reasons must be temporally neutral. Thus even if only some of the reasons generated for an agent are fully relative, S is incorrect.

But what argument is there for accepting even the modest claim that reasons *can* be fully relative? In one sense, of course, this claim is reasonably uncontroversial. Parfit's observations about relativity do bring out, I think, that there is nothing logically incoherent about a position that holds that some or all reasons are fully relative. As the discussion of P makes obvious, such reasons can be coherently described. But there is no reason for the S theorist to deny this. S need not hold that the assertion of the existence of fully relative reasons is incoherent, but only that it is *false*—that is, that there *are* no such reasons.[9] Parfit's appeal to full relativity may succeed in keeping the Present-aim theory from being dismissed out of hand as incoherent, but it does not give us any reason to prefer P over S.[10]

PAST DESIRES

Parfit's third major argument (chap. 8) focuses on what is, for many, an embarrassing commitment of the Self-interest theory. Much of the plausibility of S derives from the intuition that it cannot be rational to fail to

8. The possibility of relativizing not only to "I" and to "now" but also to "other" and to "then" calls into question the appropriateness of calling P "fully" relative. Indeed, the notion of *full* relativity may be endangered altogether.

9. For simplicity of exposition, I have avoided two qualifications in the text. First, S is troubled only by the existence of *fundamental* reasons that are fully relative; derivative reasons are no problem (cf. p. 143). Second, for S to be incorrect, no fully relative reasons need actually exist. It would be sufficient if such reasons *would* exist under the right circumstances. Such a possibility, however, requires more than the mere *logical* possibility of such reasons.

10. If one could establish that reasons *can* be fully relative—i.e., that some such reasons do exist or would under the right conditions—this would be enough to defeat S. It would not, however, be sufficient for a defense of P, which insists that all reasons *must* be fully relative. If, on the other hand, one could establish that theories must be *pure*—fully relativized or not at all—then both P and our Utilitarian version of morality would survive. Contrary to what Parfit argues (p. 148), however, I do not think that ordinary morality would survive. For it seems to me that *some* of the reasons recognized by ordinary morality are only partially relative. However, if it were established only that reasons *can* (but need not) be fully relative, then although S, P, and U would be defeated, ordinary morality would survive: it tolerantly recognizes reasons with full, with partial, and with no relativity.

give one's *future* desires direct weight. But S is equally committed to the view that one's *past* desires directly generate reasons as well, even if those desires are no longer held. As Parfit notes, this seems wildly implausible. Theory S holds that past desires should be given as much weight as current or future desires. If we cannot accept this, we have reason to reject S. And we may have reason to accept P, which, after all, instructs us to disregard desires we no longer possess.

Parfit marshals a host of cases in support of his claim that we would find it intuitively hard to believe that our past desires should be given any direct weight at all, let alone that they must be given the *same* weight as current ones. There is, unfortunately, no space here to examine these cases in detail, and it must be admitted that anyone who hopes to disarm these examples individually will have her work cut out for her.[11] But a few strategic remarks on behalf of the S-theorist are still in order.

I have presented S, P, and U as concerned with the fulfillment of various desires per se. Certain desires generate reasons to satisfy them, and this remains true even if the person who has the desire will never realize that his desire has been met. But each of the theories can also be presented in what, following Parfit, we can cáll "hedonistic" versions—theories holding that the satisfaction of desires is only important insofar as it affects the quality of the conscious experience of the person in question.

One who accepts S in its hedonistic version will not be troubled by several of Parfit's examples. Since I can no longer affect the quality of my past experiences, and since acting on a strictly past desire will not necessarily improve the quality of my current or future experience, I will not typically have a reason to act on the desire in question. Thus a hedonistic version of S can accommodate the intuition that desires that are wholly past should have no rational impact on my action.

Parfit is fully aware of the possibility of this response, and he therefore offers an ingenious series of examples (sec. 64; cf. sec. 66) in which I know that one of the following two scenarios is the case, but I am uncertain which: either I have previously undergone a great deal of physical pain, or else I will have to undergo a smaller but still significant amount of pain in the future. While I wait to discover which of these two scenarios is the truth about me, it is clear what my preference will be: although there is certainly nothing I can do to alter the facts, I will hope that the pain was in the past, even though this would mean that overall my life will ultimately contain more pain than it would were the suffering still in the future. It seems, however, that this preference cannot be accom-

11. Parfit seems especially persuasive when he stresses the irresistibility of discounting past (or future) desires that we currently take to depend on false value judgments (sec. 60). He notes the ease with which P can accommodate such cases, and he claims that S cannot. However, Parfit seems to overlook the possibility of a deliberative version of S that filters out desires based on false beliefs (false value judgments could be among these).

modated by even the hedonistic version of S, for it represents the discounting of past desires and past experiences merely on the grounds that they are *past*. Yet surely it is rationally acceptable to prefer greater past pain to admittedly lesser future pain. And so we should reject S.

The power of Parfit's examples cannot be denied. What can be denied, however, is the claim that S cannot accommodate the preference that the pain be past. Recall my earlier observation that S need not rule out any preferences as rationally unacceptable, so long as these are playing the role of ordinary n-order desires. If my preference that the pain be past is an n-order desire, then S gives it direct weight just like my other desires. Now it is true, of course, that S would object to my acting on this preference in the case where it is outweighed by my other desires—including past desires. But the cases of past versus future pains are—of necessity—cases in which I cannot act on the preference at all. Since there is no question of my actually affecting the past distribution of pains, all that becomes engaged is the impotent vocalization of the preference itself. There is nothing in the voicing of such a preference that need dismay the S-theorist.

It might be objected that although I admittedly cannot alter the past distribution of pains, it *would* be rational to act on my preference if I *could*. But I find this response unpersuasive. Even those who share the intuition behind the objection will have to admit that, for all we know, were the truths of metaphysics so radically altered that we could routinely affect the past, we might well come to hold it irrational to choose a greater past pain over a smaller future one.[12] Given that the preference cannot in fact be acted on, and since S can accept the presence of the preference itself, I believe that Parfit's cases of past and future pain do not support the rejection of S.

If I am correct, S's attitude toward the past can only be challenged with cases in which we are actually able to act—that is, cases in which the past desire can still be satisfied or frustrated (unlike the past desire not to be in pain), and we must decide whether we have a reason to fulfill that desire. As we have seen, however, hedonistic versions of S can accommodate our intuition that there is no reason to act on such past desires. So it seems to me that such a version of S cannot be successfully challenged by an intuitive appeal to cases concerning the past.[13]

But what about nonhedonistic versions of S? Note that nothing in my argument that S can accept the impotent preference that the pain

12. It is typical of the astonishing thoroughness of Parfit's discussion that he notes (p. 167) that on some accounts prayer may be a way of affecting the past. I doubt, however, whether many of us have enough faith in our intuitions about this possibility to reject the claim I make in the text.

13. One can of course still consider cases in which we must choose between a great deal of pain in the distant future and a smaller amount of pain in the near future. But I suspect that our intuitions about the rationality of such choices are likely to be either controversial or to support the claims of S.

be past turned on whether or not S is given in a hedonistic version. Thus even one who holds that it is the satisfaction of the relevant desires per se that is important should be untroubled by the cases of past and future pain. However, Parfit's original examples in which we are still able to act on wholly past desires can no longer be so readily disarmed. Advocates of nonhedonistic versions of S cannot hide behind the observation that acting on past desires cannot affect my past experience. As far as I can see, such S-theorists should insist that past desires do indeed generate reasons,[14] and they might as well also admit that this may strike us as counterintuitive. They should simply dismiss such intuitions as incorrect.

Can such a dismissal be motivated? I believe so. An S-theorist might point to the plausibility of an evolutionary account of our inclination to discount the past. A bias in favor of the present and near future was likely to aid in the survival of those who had it. If such a causal story can be told, it may explain the source of our intuitions without giving us reason to accept them as justified. Faced with Parfit's cases concerning past desires, the S-theorist can admit that our inclination to disregard such desires is natural, and deeply ingrained, while still holding that no reason has been given to believe that past desires do not generate reasons.[15] Admittedly, such an evolutionary story would not constitute a defense of S's claim that past desires *do* generate reasons. But its possibility does, I think, bring out the fundamental flaw with Parfit's third argument. In the absence of an account of the basis of a theory of rationality, rival theories can only be played off against intuitions about particular cases; and such intuitions can never be decisive.

After all, we might have hoped at the very start to establish the superiority of P over S by the mere presentation of a few cases in which it is uncontroversial that it is rational for someone to act on her current desires rather than her overall self-interest. Such a strategy would have enabled us to avoid the abstractions of Parfit's first and second arguments. But the problem with this strategy, of course, is that such cases are unlikely to be uncontroversial—at least to the committed S-theorist (although many will take Parfit's example of heroic self-sacrifice, p. 132, to be such a case). There is a deeper problem facing the strategy as well: even if we can construct cases where our intuitions strongly support P over S, it seems quite likely that we can also construct cases where our intuitions support S over P. Swapping cases is likely to bring us to an impasse. Ultimately, I believe, this is the fate of Parfit's third argument.[16]

14. Of course, an S-theorist still might try to avoid this position by appealing to some asymmetry between the metaphysical status of the past and that of the present and future. Parfit discusses such attempts in secs. 68–70.

15. When Parfit initially discusses such appeals to evolution, he seems to overlook or misunderstand the position suggested here (sec. 65, esp. pp. 169–70); later remarks, however, do consider a view similar to the one suggested (p. 186).

16. Parfit writes as though his second and third arguments concerning S would be effective against the view that all of one's desires directly generate reasons—the view that

THE NEED FOR THEORY

Parfit makes a final, closing gesture in support of P. He claims that every theory of rationality can be represented as one or another critical version of P (p. 194; cf. pp. 131, 133). Recall that critical theories can claim that agents are rationally required to possess specific desires. Thus the Self-interest theory, for example, can be represented as the critical version of P in which agents are rationally required to have as their strongest desire the bias in their own favor. Other theories of rationality could be represented, in turn, by requiring that agents have as their strongest desire the appropriate overall pattern of concern. For Parfit, then, there is no question that we should accept *some* version of P; the question is only *which* version. Restating rival theories in terms of P, he holds, will help us see more clearly what is at issue between them.

This seems to me to be doubly incorrect. Many theories of rationality, I suspect, cannot be adequately represented in terms of P.[17] But suppose I am wrong. Suppose it is true that for each theory of rationality we can find a critical version of P that gives the same verdicts as to the rationality of individual acts. Even if this is so, restating the rival theories in terms of P may actually obscure the issues instead of clarifying them. Consider S once again: the critical version of P that corresponds to S may give the

---

a reason's force "extends over time" (p. 137)—but not against the view that the bias in one's favor is "supremely rational." (See, e.g., p. 193.) This is puzzling, however. Had the appeal to full relativity succeeded, surely it would also have undermined the claim that the temporally neutral bias in one's favor is supremely rational. Similarly, such a bias would require us to act on strictly past desires, running afoul of the intuitions appealed to by Parfit's third argument. I can see no reason why the S-theorist should not always prefer to express his view in terms of the direct generation of reasons—and I have, therefore, construed S accordingly.

17. If a theory is indirectly self-defeating (see n. 5 above), then an agent whose strongest desire is to conform to the theory will not actually conform as well as he might were he to lack that desire. Such theories may demand, instead, that the agent have that set of desires and dispositions (whatever it may be) that is optimal—i.e., whose possession will lead to the greatest possible fulfillment of the theory. But this means that such a theory will not be adequately represented by a critical version of P that requires that the agent have, as his strongest desire, a desire to give weight to desires in accordance with the original theory. For example, Parfit himself argues (in chap. 1) that S is indirectly self-defeating and that it does not require agents to have the bias in their own favor as their strongest desire. But this seems to show that—contrary to Parfit's later assertions—S is not adequately represented by the critical version of P that *does* require that agents have this bias. Similar remarks are true about some versions of morality. (I owe these points to Milton Wachsberg.) It is possible, of course, to refer instead to the critical version of P that requires agents to have the particular set of desires and dispositions that turns out to be optimal for S. Such a representation, however, is not perspicuous. Since different S-theorists may disagree concerning the optimal set, each view will need to be expressed in terms of a different version of P. And these will hide the fact of the underlying common commitment to S. What is more, since S is itself *neutral* on the empirical question of what makes up the optimal set, while any given critical version of P will be essentially committed to one particular set, it is doubtful whether any such version of P will actually be a representation of S itself at all.

same individual verdicts as S, but it is still a version of P. And, as such, it holds that only *present* desires can directly generate reasons. Other desires generate reasons only indirectly—through the mediation of the rationally required bias. But the Self-interest theory *disagrees* about this fundamental point. It holds that all of one's desires *directly* generate reasons: there is no need for mediation through a currently held desire at all. As I have stressed, temporal neutrality is the fundamental issue between S and P; and this gets lost if we cast S in terms of P.

An adequate defense of the Present-aim theory will need to address directly the issue of temporal neutrality. It will need to provide an illuminating account of why temporal relativity is appropriate in a theory of rationality. In particular, it will need to explain what it is about currently held desires that allows such desires—and only such desires—to generate reasons directly. It is the absence of such an account, I believe, that dooms Parfit's defense of the Present-aim theory to failure. I do not know whether an adequate defense can be offered. But I have no doubt that one who hopes to produce such a defense cannot do better than to begin with sustained meditation upon Parfit's fascinating discussion.