

6

Rational Egoism and the Separateness of Persons

David O. Brink

Derek Parfit's wide-ranging and important book *Reasons and Persons* defends interesting and often revisionary claims about morality, rationality and personal identity in imaginative and resourceful ways. Two of the book's central themes are the nature of individual rationality and the way in which our views about it interact with our views about the nature of persons. In his discussion of these issues, Parfit examines the way in which different views of rationality treat matters of intertemporal and interpersonal distribution. He exploits parallels between these two distributional dimensions in order, among other things, to challenge rational egoism, or the 'Self-interest Theory' as he calls it. Whereas Sidgwick acknowledged a similar challenge to egoism but did not think it sound,¹ Parfit thinks that it constitutes a decisive objection to egoism. This is an important challenge. But, like Sidgwick, I believe that egoism's commitment to the separateness of persons provides a rationale for its asymmetrical treatment of intertemporal and interpersonal distribution. And because Parfit's reductionist claims about personal identity appear to challenge the separateness of persons, I shall consider his claims and argue that the egoist rationale is compatible with reductionism about personal identity.

This is an adaptation, with significant revisions, of material from my 'Sidgwick and the Rationale for Rational Egoism', in *Essays on Henry Sidgwick*, ed. B. Schultz (Cambridge University Press, New York, 1991). Some of the revisions were made during a Fellowship that I held at the Center for Advanced Study in the Behavioral Sciences, funded by an Old Dominion Fellowship from the Massachusetts Institute of Technology and by grants from the National Endowment for the Humanities (#RA-20037-88) and the Andrew W. Mellon Foundation. I would like to thank these institutions for their support. I would also like to thank audiences at Stanford University, University of California at Irvine, Massachusetts Institute of Technology, Middlebury College, and the Sidgwick conference at the University of Chicago for helpful discussion of these issues. Special thanks go to Jonathan Dancy, Stephen Darwall, Terry Irwin, Diane Jeske, Derek Parfit, Sydney Shoemaker, Alan Sidelle and Jennifer Whiting.

1 Practical Rationality: Interpersonal and Intertemporal Distribution

Practical reasoning involves assessing alternative courses of action and deciding what to do. Alternative actions will typically have different implications for the amount of benefits and harms produced and their distribution among persons and across time. If we assume that, other things being equal, the rationality of an action is proportional to the amount of value it would produce, then we can understand theories of rationality as concerned with the way in which benefits and harms rationally ought to be allocated *among persons* and *across time*.

Theories might be classified by what they say about *whose* welfare matters. On one view, only the agent's own welfare should have non-derivative rational significance for her; actions must benefit the agent in some way to give her reason for action. Although no labels seem entirely satisfactory, we might call such a view *agent-biased*. On alternative views, an agent has reason to benefit others independently of any contribution (causal or constitutive) that this makes, directly or indirectly, to her own welfare. Again, no labels seem entirely satisfactory, but let us call any view that demands concern for others that is rationally non-derivative *impartial* or *altruistic*. There are many different interpretations of impartiality. I will assume that all versions are universal in the scope of their other-regarding concern, so that the agent has some non-derivative reason to be concerned about anyone's welfare. But impartiality or altruism, as I am understanding it here, does not, as such, attach equal weight to everyone's interests. It is only a *person-neutral*, or *impersonal*, conception of impartiality that does this. On this view, an agent has equal non-derivative reason to be concerned about anyone's welfare; in deciding what to do, she should consider only the magnitude of the benefits she can produce.

Theories of rationality might also be classified by their attitudes towards the *temporal location* of benefits and harms. On one view, only present benefits should have non-derivative rational significance for the agent. We might call such a view *present-biased*. On alternative views, an agent should be *temporally impartial* in the sense that she has non-derivative reason to be concerned about benefits at any time. One form of temporal impartiality is *temporal neutrality*. On this view, an agent has equal non-derivative reason to be concerned about benefits and harms at any time; the temporal location of a benefit or harm should itself be of no rational significance to her.

Rational egoism claims that an agent has reason to do something just in so far as doing so contributes to her own overall happiness or welfare.

As such, rational egoism is agent-biased. This aspect of egoism is one way of articulating the common thought that I have special reason to be concerned about my own life and welfare. But rational egoism is not present-biased, because it assigns equal importance to benefits and harms of equal magnitude at different times in the agent's life. Indeed, egoism is not just temporally impartial; it is temporally neutral. Its temporal neutrality represents a natural articulation of the common belief that I rationally ought to be just as concerned about future goods and harms as about present ones (of equal magnitude), and that preference for a smaller present good at the cost of a greater future good is a sign of irrationality. Rational egoism is, therefore, a *hybrid* theory of rationality: it is temporally neutral but agent-biased.

Egoism can be contrasted with two *pure-bred* theories of rationality: a fully neutral theory and a fully biased theory. We might call the fully neutral theory *neutralism* or, following Sidgwick, *rational benevolence*; we might call the fully biased theory *presentism*. Neutralism is both temporally neutral and person-neutral; it holds that an agent has reason to do something just in so far as it is valuable, regardless of whom the value accrues to or when it occurs. Presentism, on the other hand, is both temporally biased and person-biased; it holds that an agent has reason to do something just in so far as it is in his own present interest.

Rational egoism and its rivals are defined *structurally*, by their attitudes to the distribution of benefits and harms across time and among persons. These theories do not themselves say what constitutes a benefit or a harm. *Subjective* theories of welfare claim that an individual's good consists in, or depends importantly upon, certain of her psychological states. For instance, *hedonism* identifies a person's good with her having pleasurable mental states or sensations. *Desire-satisfaction* theories claim that something is in an individual's interest just in case it would satisfy her actual desires or the desires that she would have in some preferred epistemic state. By contrast, *objective* theories of welfare identify a person's welfare or happiness with certain character traits, the exercise of certain capacities, and the possession of certain relationships to others and the world, and claim that these things are good for the agent independently of their producing pleasure for her or being the object of her desire. Like Parfit, I want to focus on a structural worry about egoism's hybrid character. For this reason, wherever possible, I shall abstract from the different versions of egoism and its rivals that result from incorporating different evaluative assumptions.

Because rational egoism has a hybrid structure, it may seem arbitrary. It may seem that we should distribute goods and harms by the same principles within lives and across lives. In *The Methods of Ethics* Sidgwick raises, but ultimately rejects, this parity argument as a possible

difficulty for egoism (ME, pp. 418–19). In *The Possibility of Altruism* Thomas Nagel makes a similar parity argument on behalf of altruism.² In chapter 7 of *Reasons and Persons* Parfit argues that Sidgwick underestimated the power of the parity argument. Like Nagel, he accepts parity and rejects egoism, but he draws different conclusions. Though Parfit is sometimes content with the parity claim that is agnostic between the fully neutral and fully biased pure-breds, he seems to think that the natural conclusion of the parity argument is fully biased presentism. For his own 'Present-aim Theory' is fully biased, or, as he says, fully 'relative'; he takes it to be the main challenge to egoism (p. 117), and he treats the parity argument as part of the 'Appeal to Full Relativity'.

2 Parfit's Assumptions

Parfit describes his own rival theories in somewhat different terms than those I use for mine. His three rivals are the fully relative Present-aim Theory, P, the hybrid Self-interest Theory, S, and the fully neutral morality, or Neutralism, N. P tells each agent to aim at what will best achieve his present aims (p. 92). S tells each agent to aim at what will make his life as a whole go as well as possible (p. 3). And N seems to tell each to aim at the best outcome overall (pp. 94–5, 126, 138–44). Parfit also distinguishes between hedonistic, desire-fulfilment and objective list theories of value (p. 4).³ Because I shall formulate and assess Parfit's arguments in terms of my distinctions, I should comment briefly on ways in which our assumptions differ.

First, we are both concerned with the way in which theories of practical rationality distribute benefits and harms across persons and across time. But whereas Parfit contrasts relative and neutral distribution, I contrast biased and neutral distribution, and see neutral distribution as a special case of impartial distribution. Though I shall focus on neutral forms of impartiality, I think that it is important to recognize, as Sidgwick and Nagel do, that perhaps the most significant fact about the egoist view of interpersonal distribution is that it rejects impartiality; it recognizes only derivative reason to benefit others. This fact also explains why I speak of temporal or personal bias, rather than relativity. As agent relativity is generally understood, a reason is agent-relative if its general form involves essential reference to the person who has it.⁴ But many theories besides S and P are agent-relative, including many that recognize non-derivative reason to benefit others. For example, C. D. Broad sketched a theory that he called 'self-referential altruism', according to which one has non-derivative reason to benefit others, but the weight or strength of one's other-regarding reasons is a function of the nature of

the relationship in which one stands to the beneficiary.⁵ Self-referential altruism is agent-relative, but it is impartial or altruistic because it recognizes non-derivative reason to benefit others. If the most significant aspect of the egoist's account of interpersonal distribution is that it does not recognize non-derivative reason to benefit others, then it is misleading to focus on agent relativity here, as Parfit does, because this puts some impartial or altruistic theories, such as self-referential altruism, in the same camp as egoism, rather than in the same camp as neutralism. If so, the focus here should be on what I call agent bias – the denial of non-derivative concern for others – rather than agent relativity.

Second, I construe all three structural rivals as theories of rationality. Unlike Parfit, who sometimes construes the fully neutral theory as morality, I want to treat it as a theory of rationality, just like the other two theories; this makes their rivalry clearer, and allows us to assess the parity claim in a more straightforward manner. This will be clearest if we think of the fully neutral theory, as Sidgwick did, as rational benevolence; it stands to rationality as utilitarianism or consequentialism stands to morality.

Third, Parfit's P is in some ways different from presentism. P claims that one has reason to do something just in so far as it would satisfy one's present desires or aims; whereas presentism claims that one has reason to do something just in so far as it would promote one's present interests or happiness. I think there are two different ways to understand P and its relation to presentism.

On one reading, P's focus on present aims, rather than present interests, of the agent is very significant. On this view, P is not primarily a view about how to decide among beneficial and harmful alternatives; rather, it reflects assumptions about the purely instrumental character of practical rationality, according to which action is rationally assessable only in so far as it is instrumental to the satisfaction of the agent's desires or aims. But this Humean, or instrumental, conception of P is not a conception of the intertemporal and interpersonal distribution of benefits and harms.⁶ Interesting as this conception of P is, it is not a good structural rival to egoism.⁷

Alternatively, we might not regard P's focus on present aims, rather than present interests, of the agent as very significant. For we can understand P as a special case of presentism: namely, presentism with a desire-satisfaction theory of value. But in the present context we should focus on presentism, rather than its special case P. This is because the parity argument aims to compare three theories that abstract from different possible claims about value and make competing claims about intertemporal and interpersonal distribution. This is how egoism and neutralism are formulated; the third member of the triad should also

abstract from claims about value and make claims about intertemporal and interpersonal distribution. If so, the fully biased theory should be presentism, not P.⁸

These are my main reasons for formulating the structural parity issue as I do. The differences between Parfit's formulation and my own should cause neither confusion nor distortion.

3 The Parity of Intertemporal and Interpersonal Distribution?

Unlike neutralism or presentism, rational egoism is a hybrid theory of rationality; it is temporally neutral and agent-biased. Can this hybrid character be justified? Is it reasonable to treat time and person differently, as egoism does? Or is egoism an unhappy compromise between the two pure-breds? Egoism is committed to saying that it makes all the difference *on whom* a benefit or burden falls, and none whatsoever *when* it falls. On reflection, this may seem arbitrary. As Sidgwick notes, in discussing the 'proof of utilitarianism',

I do not see why the axiom of Prudence [rational egoism] should not be questioned, when it conflicts with present inclination, on a ground similar to that on which Egoists refuse to admit the axiom of Rational Benevolence. If the Utilitarian [neutralist] has to answer the question, 'Why should I sacrifice my own happiness for the greater happiness of another?' it must surely be admissible to ask the Egoist, 'Why should I sacrifice a present pleasure for a greater one in the future? Why should I concern myself about my own future feelings any more than about the feelings of other persons'. (ME, p. 418)

The egoist asks the neutralist, 'Why should I sacrifice my own good for the good of another?'. The egoist doubts that concern for others is non-derivatively rational. But the presentist can ask the egoist, 'Why should I sacrifice a present good for myself for the sake of a future good for myself?'. The presentist doubts that concern for one's future is non-derivatively rational. These doubts may seem parallel. We must decide where among lives and when within lives to locate goods and harms. Because both are matters of position or location, we may think that they should be treated in the same way. Parfit does (p. 140). He writes:

As a hybrid, S can be attacked from both directions. And what S claims against one rival may be turned against it by the other. In rejecting Neutralism, a Self-interest Theorist must claim that a reason may have force only for the agent. But the grounds for this claim support a further claim.

If a reason can have force only for the agent, it can have force for the agent only at the time of acting. The Self-interest Theorist must reject this claim. He must attack the notion of a time-relative reason. But arguments to show that reasons must be temporally neutral, thus refuting the Present-aim Theory, may also show that reasons must be neutral between different people, thus refuting the Self-interest Theory. (p. 144)

If present sacrifice for future benefit is rational, why isn't sacrifice of one person's good for the sake of another's? In this way, the appeal to parity may support neutralism. This is roughly Nagel's view.⁹ His primary aim is to argue against egoism's agent bias and in favour of impartiality, or altruism, and he relies on the parity of intertemporal and interpersonal distribution to do so. Just as the interests of an agent's future self provide him with reasons for action now, so too, Nagel argues, the interests of others can provide him with reasons for action. Failure to recognize temporal neutrality involves temporal dissociation – failure to see the present as just one time among others – and failure to recognize impartiality, or altruism, involves personal dissociation – failure to recognize oneself as just one person among others.¹⁰

Nagel's remarks about the 'combinatorial problem' show that he is sceptical of an impersonal interpretation of impartiality.¹¹ None the less, his appeal to parity seems to require neutralism and not just impartiality. For he appeals to parity to argue from egoism's temporal neutrality to non-derivative concern for others. But if intertemporal and interpersonal distribution must be isomorphic, and we accept a temporally neutral interpretation of intertemporal impartiality, then we seem forced to accept a person-neutral interpretation of interpersonal impartiality.

Alternatively, we might treat time and person as parallel and argue from the agent bias that egoism concedes to temporal bias, in particular, present-bias. If my sacrifice for another is not rationally required, it may seem that we cannot demand a sacrifice of my current interests for the sake of distant future ones. If so, we will think that it is only the present interests of the agent that provide her with non-derivative reason for action. Though Parfit mentions Nagel's fully neutral response to parity, it is the fully biased response that he develops and thinks Sidgwick anticipated (pp. 137–44).

4 Egoism and the Separateness of Persons

Whereas Parfit thinks that egoism cannot meet this challenge, Sidgwick thinks this challenge is unanswerable only on a very extreme kind of scepticism about personal identity. The passage in which Sidgwick raises the parity theme continues this way:

It undoubtedly seems to Common Sense paradoxical to ask for a reason why one should seek one's own happiness on the whole; but I do not see how the demand can be repudiated as absurd by those who adopt the views of the extreme empirical school of psychologists. . . . Grant that the Ego is merely a system of coherent phenomena, that the permanent identical 'I' is not a fact but a fiction, as Hume and his followers maintain; why, then, should one part of the series of feelings into which the Ego is resolved be concerned with another part of the same series, any more than with any other series? (ME, pp. 418–19)

Sidgwick sees the presentist's challenge to egoism, but thinks we can meet this challenge, because it depends upon denying that persons persist. Later, he suggests that the *separateness of persons* is what underlies rational egoism.

It would be contrary to Common Sense to deny that the distinction between any one individual and any other is real and fundamental, and that consequently 'I' am concerned with the quality of my existence as an individual in a sense, fundamentally important, in which I am not concerned with the quality of the existence of other individuals: and this being so, I do not see how it can be proved that this distinction is not to be taken as fundamental in determining the ultimate end of rational action for an individual. (ME, p. 498)

In these two passages Sidgwick appeals to the separateness of persons to justify egoism. I think he must suppose that the separateness of persons has both *metaphysical* and *normative* aspects. On the one hand, the separateness of persons involves recognition of ourselves as numerically distinct and temporally extended entities. But Sidgwick seems to think that this metaphysical claim commits us to normative claims as well. The second passage suggests that to recognize myself as distinct from others and temporally extended requires me to adopt patterns of concern that exhibit a bias towards myself. On one reading, the bias is supposed to be legitimated by prior recognition of the separateness of persons. But this seems to leave the process of legitimation unexplained. Alternatively, we might understand Sidgwick to be suggesting that the bias is part of what constitutes the separateness of persons. Though Sidgwick does not develop this suggestion, it seems not unreasonable.

It might be claimed that part of what it is for me to be a distinct person who persists through time is for me to form and act from a particular perspective on the world. Part of having and acting from a particular perspective involves forming and acting on intentions and goals in ways that display self-concern. Some of this self-concern involves backward-looking displays of trust. I form my intentions and goals as the result of

my own previous deliberations. Though I can take account of other people's opinions in my own deliberations, forming my own intentions involves accepting and endorsing my own previous deliberations, even when I have reason to suspect that they have been imperfect and that the deliberations of others are likely to have been better. Other parts of this self-concern are forward-looking. Whereas some of my desires are impersonal in nature, many of them are personal. For instance, if I am trying to prove a certain theorem, not only do I want the theorem to be proved, but also I want to be the one to prove it, and this personal desire structures my plans and activities. We can see that my personal desire is not purely instrumental to the satisfaction of my impersonal desire by noting that my aims would not be fully accomplished if someone else were to prove the theorem, even if her proof were superior to the one I would have offered.¹² It seems an important part of what makes people distinct that they form and act on such personal desires.¹³ Indeed, if people's activities were not governed for the most part by personal desires, it is hard to see how there could be impersonal desires that valuable lives be led. What does a valuable life consist in if not, in large measure, the satisfaction of suitable personal desires?¹⁴ Intentions are like personal desires, because one always figures in one's own intentions. I always intend to bring something about through my own actions. For instance, I may want my children to be raised well, but I intend to do things to bring this about, typically by raising them myself. And this intention is not satisfied if others take my children, without my consent, and raise them well, even if they raise them better than I would have. And, of course, having such self-involved intentions gives rise to a generalized concern for one's own future (e.g. anticipation, fantasy and anxiety) and further self-involved intentions as a way of ensuring that one will be able fulfil these original intentions. Moreover, in adopting intentions, I treat the object of my intentions and subsidiary objectives as more or less fixed in ways that constrain my future deliberations.¹⁵ For example, if I intend to finish my article this coming summer, this constrains deliberation about the scheduling of the family vacation and about whether to accept other professional commitments. But then, when I adopt self-involved intentions, the associated self-concern must also constrain my deliberations in ways that preclude seeing myself merely as a resource for the production of impersonal value.

Another part of what it is for me to be a person with a particular perspective on the world is that I am located at a particular point in a network of interpersonal relationships and commitments, even if the place I occupy in this interpersonal network changes over time. I form friendships, marry and raise children, and these relationships help define my life. Part of standing in such relationships to intimates is for me to

make them the object of my personal projects. I want to do certain things with them and for them, and I value these activities out of proportion to their impersonal value. In fact, my concern for those to whom I stand in special relationships shapes my conception of my own interests and the standards by which I judge my life to be a success.

If so, part of what it is for me to be a distinct, temporally extended person is for me to have a particular perspective on the world that displays a concern for my past and future self that is not proportional to the impersonal value of my activities. But then part of what it is for me to be a separate person is for me to be unwilling to sacrifice my interests without appropriate compensation. This would be one way of explaining the links Sidgwick seems to see between the separateness of persons and demands that sacrifices be compensated (*ME*, pp. 171, 174, 404, 498, 499, 501, 502).¹⁶

This appeal to the separateness of persons provides a rationale for egoism's hybrid character. The metaphysical separateness of persons plays two roles in this rationale. First, it is supposed to explain the appropriateness of the normative claim that it is unreasonable to make uncompensated sacrifices. Second, this compensation principle and the metaphysical separateness of persons explain the asymmetry between intrapersonal and interpersonal distribution. For there is automatic *intrapersonal* compensation, but no automatic *interpersonal* compensation. Compensation requires that benefactors also be beneficiaries, and for compensation to be automatic, benefactor and beneficiary must be one and the same. In the diachronic, intrapersonal case, one's sacrifice of a present good for a (greater) future good is rational, because there is compensation later for the earlier sacrifice; benefactor and beneficiary are the same. This explains temporal neutrality. But in the interpersonal case, benefactor and beneficiary are different people; unless the beneficiary reciprocates in some way, the agent's sacrifice will be uncompensated. This explains agent bias.

Before we discuss the adequacy of this rationale, we should perhaps explain some implications of temporal neutrality and agent bias.

5 Temporal Neutrality

Because diachronic, intrapersonal compensation is automatic, it is the importance of the agent's interests, and not their temporal location, that determines what an agent has most reason to do. As Sidgwick claims:

Hereafter *as such* is to be regarded neither less nor more than Now. It is not, of course, meant, that the good of the present may not reasonably be

preferred to that of the future on account of its greater certainty; or again, that a week ten years hence may not be more important to us than a week now, through an increase in our means or capacities of happiness. All that the principle affirms is that mere difference of priority and posteriority in time is not a reasonable ground for having more regard to the consciousness of one moment than to that of another. (ME, p. 381; cf. pp. 111–12, 124n.)

As Sidgwick notes, this kind of temporal neutrality is perfectly compatible with the claim that rational *planning* should take into account the uncertainty of my continued existence. We can explain this point by drawing a useful distinction between *subjective* and *objective* rationality (cf. p. 153; ME, pp. 207–8, 394–5). Claims of objective rationality are claims about what an agent has reason to do given the facts of a situation, whether he is aware of these facts or in a position to recognize the reasons that they support. Claims of subjective rationality are claims about what the agent has reason to do given his beliefs about his situation or what it would be reasonable for him to believe about his situation. The egoist can admit that the existence of my near future is more certain than the existence of my distant future, and that this epistemic fact should affect what it is subjectively rational for me to do; he claims only that in so far as I have both present and future interests, they provide me with equally strong objective reasons for action.

Temporal neutrality claims that what I now have most reason to do is what will most advance my overall good. This may require present sacrifices for my future benefit (now-for-then sacrifice). In one sense, this requires acting on interests that I do not now have, but this does not require me to act as if I now had those interests. For instance, it does not require me to eat or buy foodstuffs now in order to satisfy my future hunger. Where temporal neutrality requires now-for-then sacrifice, it requires only that I act now so as to be able to secure my later good at that time. For instance, it requires only that I set aside or employ my resources now so that I will then have the means available to satisfy the hunger that I will have later.¹⁷

6 Agent Bias

Because I am a separate entity, I am not automatically compensated by benefits to another, and this makes interpersonal sacrifice problematic. But compensation does not imply that others cannot be beneficiaries of the agent's actions; it requires only that the agent be an appropriate beneficiary. For instance, agents will often receive benefits in return for those they confer, as in most systems of co-operation for mutual advan-

tage. In this way, the egoist can often justify being a reliable benefactor (cf. ME, pp. 164–70, 499–503).

The egoist may also be able to justify other-regarding action and concern as a kind of 'false target' for promoting the agent's own good. Sidgwick insists on the so-called paradox of egoism, according to which an agent may best promote her own interest not by doing things with the aim of maximizing her own welfare, but by developing concerns for people and activities for their own sakes (cf. ME, pp. 48, 136, 403). The paradox of egoism is supposed to explain why even hedonistic egoists should care for others for their own sakes.

Finally, versions of rational egoism that employ an objective conception of welfare have additional resources for justifying other-regarding conduct and concern. Many objective conceptions of welfare recognize a variety of social or other-regarding components in a person's good. They claim that family relationships, friendships and social relationships involving mutual concern and respect make our lives more valuable than they would otherwise be. By having friends and co-operating with others on a footing of mutual concern and commitment, we are able to exercise new capacities, secure ourselves against a variety of misfortunes, and generally extend our interests in new ways. On such views, the good of others is *part of my own good*; if so, benefits to them are benefits to me. On such a view, my reason to benefit others is derivative from my reason to benefit myself, but my concern for others is not purely instrumental. If such a view can be defended, it will allow for the possibility of interpersonal compensation, and so provide the basis for a robust defence of other-regarding conduct and concern on egoist grounds.

Of course, it is an open question exactly how far such strategies can be used to provide agent-biased justification of other-regarding conduct and concern. My immediate purpose is only to forestall possible misunderstandings of the commitments of agent bias and to insist that the success of such strategies is a genuinely open question.¹⁸

7 The Egoist Rationale

According to Sidgwick, the separateness of persons explains why the egoist holds a hybrid theory of rationality consisting of agent bias and temporal neutrality. The egoist will see no reason to follow the neutralist, who infers person neutrality from temporal neutrality. Stages of other people are not stages of me, present or future, and as a result, I may not be compensated for my sacrifices to others. The egoist will see no reason to follow the presentist, who infers temporal bias from agent bias. Unlike the stages of other people, my future self is a stage of me, and therefore

I do receive compensation when I make present sacrifices for my own future benefit. Diachronic, intrapersonal compensation is automatic; interpersonal compensation is not.

Because the egoist offers a rationale for rejecting parity, egoism's hybrid character cannot be represented as arbitrary or unmotivated. Indeed, on this rationale, whereas there is a kind of asymmetry in the egoist's treatment of intrapersonal and interpersonal distribution, this asymmetry reflects a deeper, symmetrical treatment of these two distributional dimensions. Though egoism is intertemporally neutral but interpersonally biased, both commitments follow from the separateness of persons and, in particular, the compensation principle. Egoism applies the compensation principle to both intrapersonal and interpersonal contexts: because there is automatic intrapersonal compensation, egoism is temporally neutral; because there is no automatic interpersonal compensation, it is personally biased.

The compensation principle is not uncontroversial. Its most controversial aspect, I believe, is its claim that uncompensated interpersonal sacrifice is unreasonable. But neither is this claim absurd.

Sidgwick motivates the compensation by appeal to the separateness of persons. We can explain the link, I have suggested, if the compensation principle expresses a form of self-concern that is itself constitutive of the separateness of persons. Even if we do not suppose that special concern is itself constitutive of the separateness of persons, it is commonly believed that we do have special reason to be concerned about our own lives that we don't have with regard to the lives of others. The rationality of an agent's actions seems not in general to be proportional to the good that she does. The compensation principle provides one explanation of why such special concern is appropriate.

Moreover, the compensation principle provides a natural explanation of traditional worries about the rational authority of morality. The basis of this worry is the claim that morality requires agents to benefit or respect others in ways that seem to constrain the pursuit of their own interest or happiness. But in supposing that there is a difficulty in defending the rational authority of other-regarding moral demands, we seem to assume that the rationality of sacrifice requires compensation, as the compensation principle claims.

It might be said that these claims do not require anything as strong as the compensation principle or egoism's agent bias. For the compensation principle implies that I have no reason to make an uncompensated sacrifice for another, no matter how small my sacrifice and how large the benefit it would confer on another, and agent bias implies that I have no non-derivative reason to benefit others. By contrast, it might be claimed, special concern and traditional worries about the authority of morality

could be explained by weaker assumptions. We might not require full compensation for all sacrifices; we might not require compensation for small sacrifices, and we might apportion compensation, where we give it, in light of the seriousness of the sacrifice and the amount of benefit conferred by it. This might result in a more subtle kind of personal bias. On one such view, agents have non-derivative reason to benefit others, but, all else being equal, their self-regarding reasons are weightier than their other-regarding ones. By our definitions, such a theory would be impartial or altruistic, but it would assign weight to reasons according to an *agent multiplier*. In fact, the resulting theory might resemble Broad's self-referential altruism, construed here as a theory of rationality. These weaker assumptions would still allow us to make sense of special concern and raise worries about the authority of morality, provided morality contains more neutralist elements. Moreover, something like self-referential altruism might seem like an intuitively attractive compromise between agent bias and person neutrality.

But this apparent appeal is a source of trouble in the present context. For the structural objection to egoism assumes the parity of intrapersonal and interpersonal distribution; parity, it is assumed, requires a pure-bred conception of rationality. But self-referential altruism is neither agent-biased nor person-neutral (it is altruistic with an agent multiplier), much less fully biased or fully neutral. It is itself hybrid, with a vengeance, and so can provide no comfort to the friends of parity and pure-breds.

Indeed, if parity is a virtue, then the egoist can exploit parity to argue against self-referential altruism. As I have indicated, egoism respects parity at a deep level, because it justifies agent bias and temporal neutrality by applying the compensation principle to both intrapersonal and interpersonal contexts. Paradigmatic forms of rational deliberation involve deciding to take some course of action that one would otherwise view as indifferent or harmful as a means or necessary condition to doing something else that is best overall. This familiar sort of now-for-then sacrifice is rational precisely because the agent is suitably compensated later for her present sacrifice. Indeed, intrapersonal sacrifice seems justified if and only if it is adequately compensated. This is what explains temporal neutrality. But the very compensation principle that justifies temporal neutrality in intrapersonal contexts justifies agent bias in interpersonal contexts. If we believe that intrapersonal and interpersonal distribution must be regulated by the same principles, then this very familiar and attractive explanation of temporal neutrality requires us to reject self-referential altruism.

Moreover, some versions of rational egoism can help themselves to whatever appeal self-referential altruism might have. As I will explain

below (in section 16), rational egoism can justify concern for others, on the agent's part, proportional to the strength of the psychological relations in which they stand to him, by representing such relations as extending the agent's own interests.¹⁹ If so, we may be able to represent the central claims of self-referential altruism as theorems of rational egoism.

Finally, it is important to notice that the egoist's opponent in the parity argument relies on a compensation principle. The parity argument treats the egoist's demand, 'Why should I sacrifice my interests for another?', and the presentist's demand, 'Why should I sacrifice my present interest for my future interests?', as parallel. So Parfit's attempt to argue from egoism's agent bias to presentism's full bias appeals to the need for compensation, because egoism challenges neutralism by appealing to the need for compensation. Moreover, Nagel seems to assume something like the compensation principle when, in discussing the combinatorial problem, he rejects the interpersonal balancing that a person-neutral theory allows by appeal to 'the extremely strict position that there can be no *interpersonal* compensation of sacrifice'.²⁰ Compensation is a controversial demand, but it is not the issue in this structural debate.

The egoist rationale also rests on the metaphysical separateness of persons. Sidgwick thinks that these metaphysical assumptions about personal identity are seriously challenged only by Humean views about personal identity. And Parfit claims that his own reductionist view of personal identity subverts egoism. Later (sections 9–16), I will discuss whether this egoist rationale is compromised by reductionist views of personal identity. But Parfit explicitly claims that the parity argument does not depend upon any special assumptions about personal identity (p. 139). This I deny. As long as agents are metaphysically distinct, interpersonal compensation is problematic; as long as agents are temporally extended, diachronic, intrapersonal compensation is automatic.

My purpose here is not to provide a full-blown defence of rational egoism or to argue that its pure-bred rivals are untenable. In particular, I have not claimed that the compensation principle is unproblematic or that the implications of agent bias and temporal neutrality are fully acceptable.²¹ What I have argued is that there is a principled rationale for the hybrid character of egoism; indeed, egoism respects parity in so far as it regulates intrapersonal and interpersonal distribution by the same compensation principle.

8 Temporal Neutrality and the Bearers of Reasons

The presentist might claim that we have not appreciated the force of his appeal to full bias or relativity. The egoist insists that intrapersonal

compensation is automatic, because benefactor and beneficiary are the same. This assumes that it is the temporally extended being – the person – that is the agent and the bearer of reasons for action. But the presentist might aim to challenge this natural assumption. The egoist is being asked to explain why the interests of my present self should be sacrificed for the benefit of my future self. This may make it look as if it is a person-stage, *me-now*, rather than me, who is the bearer of reasons. If so, the egoist's rationale would seem to fail. *Me-now* can fairly complain that it is being asked to make an uncompensated sacrifice. After all, the benefits of its sacrifice accrue not to *me-now* but to *me-later*. *Me-now* seems no more compensated for its sacrifice to *me-later* than I am by my sacrifice to you. Perhaps the compensation principle forces us to reject egoism in favour of presentism.

The issue seems to depend on whether we think of the agent and the bearer of reasons as the temporally extended person or as a person-stage. There is automatic intrapersonal compensation, but interstage compensation is problematic. We may ask the egoist whether *my present self* (*me-now*) has reason to sacrifice its interests for those of my future self (*me-later*). Or we may ask instead, as Sidgwick and Parfit both do, whether *I* have reason to sacrifice my current interests for the sake of my future interests.²² The egoist can answer the second question; I am compensated for a present sacrifice by a (greater) future gain, because that future self is a part of me. Temporal neutrality seems threatened if and only if we formulate the challenge this first way. Though Parfit makes the appeal to full relativity or bias, this appeal to person-stages is not part of his view (pp. 95, 135, 139–45). None the less, it provides a defence of presentism that is worth considering.

We normally suppose that persons are agents and the bearers of reasons. Talk about person-stages may seem to be merely a convenient way of talking about persons at particular times. If so, egoism's temporal neutrality is secure, for there is automatic intrapersonal compensation. But even if we agree that person-stages are entities in their own right, perhaps four-dimensional entities,²³ is there any reason to think that they are agents?

It might be said that practical reasoning is necessarily temporally indexed. My deliberations are always about what I should do, plan, omit or postpone *now*, and I must deliberate on the basis of my *current* beliefs, aims and values.²⁴ But this does not show that agency must be understood in terms of person-stages, rather than persons. For as I have described this aspect of practical reasoning, it connects deliberation and action with what is true of agents at particular times; it does not involve commitment to person-stages.

Indeed, the suggestion that agents are person-stages is problematic.

not just unmotivated. Its plausibility depends upon the 'life-span' of a person-stage. If the appeal to temporal bias is to be part of an appeal to full bias, and if this appeal is supposed to show that person-stages, rather than persons, are agents, then person-stages must be understood as time-slices or temporally minimal cross-sections of a person's life. Otherwise, the appeal to bias would not be full or complete. Call such a temporally minimal person-stage a *person-slice*.

It is difficult to regard person-slices as agents. For one thing, this conception of agency requires that person-slices have interests. If me-now is to be an agent concerned to advance its own interests, me-now must have interests. But it is not clear that person-slices do have interests (cf. p. 135). Whether the entity me-now can have interests depends upon which theory of welfare is correct.

If pleasure is a simple, qualitative sensation or mental state and a hedonistic theory of welfare is correct, then perhaps me-now has interests. Person-slices may contain qualitative mental states, such as pleasure. However, this is not the only version of hedonism. On a more plausible version, pleasure and pain are functional states: pleasure is a mental state or sensation such that the person having it wants it to continue and will, *ceteris paribus*, undertake actions so as to prolong it, whereas pain is a mental state or sensation such that the person having it wants it to cease and will, *ceteris paribus*, take action to make it stop (cf. p. 493; ME, pp. 42–3, 46, 127, 131, 402). But according to this version of hedonism, pleasure is already a state that is essentially the state of a temporally extended entity. If so, the bearer of pleasures cannot be a person-slice.

Moreover, hedonism seems an implausible theory of welfare, because a large part of a person's good seems to consist in his *being a certain sort of person* – that is, a person with a certain sort of character who exercises certain capacities and develops certain kinds of personal and social relationships. This is true not only on objective theories of value, which take welfare to consist in being a certain sort of person, but also on desire-satisfaction theories of welfare. Our desires are the desires of temporally extended beings, and the content of most of our desires is conditioned by this fact. If so, our conception of interests and welfare (i.e. of what makes a life go well) is fundamentally *diachronic*. And this implies that it is temporally extended beings, as the egoist assumes, rather than person-slices, who are the bearers of interests.

Not only is it doubtful whether person-slices have interests, it is also questionable whether having interests is sufficient for having reasons for action. A person-slice will not persist long enough to perform actions or receive the benefits of actions. If so, then person-slices cannot have reasons for action even if it is possible for them to have interests. Indeed,

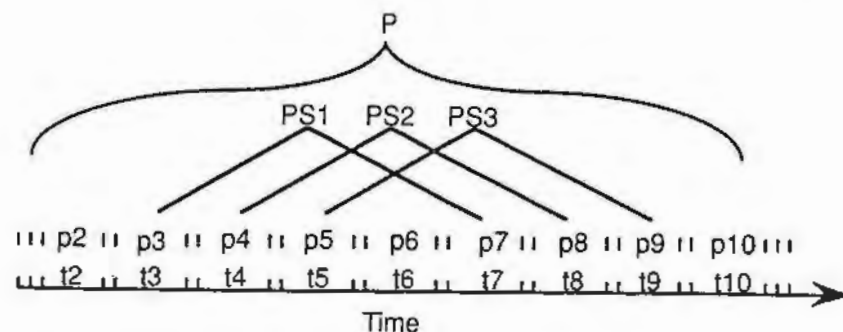


Figure 6.1 Overlapping temporal parts

more generally, as we have noted, a paradigmatic form of intrapersonal practical reasoning involves deciding to make short-term sacrifices as a means or necessary condition of achieving valuable goals. The rationale for such now-for-then sacrifice is that it is compensated; sacrificer and beneficiary are the same. The presentist also appeals to the compensation principle. But if agents are conceived of as person-slices, then we must reject one of our clearest paradigms of practical reason as irrational. Indeed, it is unclear what scope this view leaves for practical reasoning.

What if we construe person-stages, not as person-slices, but as temporally extended, cross-sections of a person's life? Call such a person-stage a *person-segment*. Person-segments, unlike person-slices, can have some interests and can live long enough to perform and receive some of the benefits of actions. The appeal to full bias, so construed, would also seem to challenge rational egoism's temporal neutrality, because me-now, conceived as a person-segment, may not be compensated for its sacrifice to me-later, conceived of as a later segment of the same person.

However, appeal to person-segments leads to a proliferation of agents and consequent indeterminacies in questions about practical deliberation that are troubling and unnecessary. Person-segments are made up of temporally related series of shorter person-stages (e.g. slices) and so contain parts that also belong to other person-segments; that is, person segments overlap. Consider a person consisting of person-stages, including stages p2 through p10 (figure 6.1). The person segment PS1 consisting of person-stages p3 through p7 contains within it, for example, part of the person-segment PS2, consisting of p4 through p8, and part of the person segment PS3, consisting of p5 through p9. Suppose the time is t6. If, as we normally assume, it is the person P that is the subject of practical deliberation, the egoist rationale allows us to identify her reasons for action. If person-segments are agents, whose reasons for action should

we be concerned about now: those of PS1, those of PS2, or those of PS3? (This names only three of infinitely many possibilities.) Of course, we can identify the reasons for action of each of these entities. But reasons for action are supposed to be related in some way to practical deliberation. Whose practical deliberation is in question?

The friend of person-segments might reply by arguing that some person-segments are more *salient* than others. Person-segments are individuated temporally, but some display more internal psychological coherence and integrity than others. Some segments of a person's life cut that life at its psychological joints, whereas others do not. So, it might be argued, at any point in time, the subject of practical deliberation should be identified with the salient person-segment existing at that time.

This suggestion reduces, but does not eliminate, indeterminacy. This is because salient segments will themselves overlap. Psychological change is generally fairly continuous, but proceeds at different rates along different dimensions of someone's personality. Even if there were mutual interaction among all aspects of a person's psychological make-up, there would still be distinguishable cognitive and affective aspects of a person's mental life whose development and careers would be staggered (cf. linguistic skills, mathematical skills, athletic skills, political beliefs, tastes in food, friendships, and career goals and expectations). This is just to say there are a great many different psychological joints at which to cut for salience. If so, there will be overlapping salient segments, producing significant, even if less extreme, proliferation and indeterminacy. Indeed, so long as the person is herself a salient segment, at any time there must be at least two possible agents on the person-segment view.

Moreover, these person-segments, salient or otherwise, do and must interact and co-operate. They do interact and co-operate, much as distinct individuals interact and co-operate in groups, in order to plan and execute long-term projects and goals. They must interact and co-operate if only because they have to share a body and its capacities in order to execute their individual and collective goals, much in the way that individuals must sometimes interact and co-operate if they are to use scarce resources to mutual advantage.²⁵ Indeed, both the ease and the necessity of interaction among person-segments will be greater than that among persons, because the physical constraints and the reliability of fellow co-operators are greater in the intrapersonal case. But this means that person-segments will overlap with each other; they will stand to each other and the person much as strands of a rope stand to each other and the rope. Though we can recognize the overlapping strands as entities, the most salient entity is the rope itself. So, too, the most salient entity is

the person, even if we can recognize the overlapping person-segments that make up the person.

In this way, person-segments represent a rather arbitrary stopping place. If the appeal to full bias argues for agents with shorter life-spans than persons, then an appeal to *full* bias ought to argue for person-slices as agents. But if, as I have argued, that conception cannot be maintained, then it seems arbitrary to settle on person-segments. Once we extend the life-span of the agent beyond that of a person-slice, it seems that we should keep going until we reach the entity with the most natural borders: namely, the person.

These appear to be reasons for preserving the normal assumption that it is persons that are agents. But is this assumption itself incoherent? I have identified the person with a temporally extended entity, some of whose parts lie in the future. But then the person is in one sense 'not all there' at the time of deliberation and action. How then could the person be the agent who deliberates and acts and possesses reasons for action?

This raises difficult issues, but I doubt that they threaten the assumption that it is persons that are agents.²⁶ Notice, first, that person-slices seem to be the only candidates for agency that avoid some form of this objection. For person-segments extend from the instant of deliberation or action into either the future or the past (or both); so person-segments are also entities with parts that are 'not all there' at the time of deliberation or action. Only one person-slice is 'all there' at this time. But we have already seen that such a conception of agency is indefensible. We might, therefore, wonder whether the agent or entity whose interests determine what rationally ought to be done need be 'all there' at the time of action.

Consider an analogy with nations. We speak of nations as actors that enact legislation, start wars, and so on. We also think of nations as having interests and acting in their interests. But a nation is composed, at least in part, by its entire current population. And there is certainly some sense in which the entire population does not enact legislation or start wars. Instead, certain individuals or groups act as representatives of a larger, spatially dispersed group of which they are members. We do not conclude that nations cannot be actors or the bearers of interests. Instead, we conclude that a nation can act when its deputies act on behalf of the national interest – that is, the interest of the spatially dispersed group. Similarly, the present self can act as representative of the temporally dispersed entity, the person, by acting in the interest of this being. If so, then the fact that the temporally extended person is 'not all there' at the time of action is not a reason to deny that it is the actor or the entity whose interests determine what agents have reason to do. On this

assumption, there is automatic diachronic, intrapersonal compensation, and so compensation does justify temporal neutrality.

9 Rational Egoism and Reductionism about Personal Identity

As we have seen, Sidgwick thinks that rational egoism appeals to the separateness of persons, and that it would be seriously challenged by a Humean account of personal identity (*ME*, pp. 418–19). Nor is such linkage between personal identity and our attitudes towards self-concern and future concern uncommon. Butler claims that special concern for oneself and moral responsibility would be undermined by Locke's account of personal identity in terms of memory connectedness.²⁷ In part 3 of *Reasons and Persons* Parfit likens his own reductionist account of personal identity to the views of Locke and Hume, and argues, among other things, that reductionism undermines egoism's temporal neutrality (pp. 139, 307–20). Because my concern is whether reductionist accounts of personal identity undermine egoism's hybrid rationale, I shall assume that Parfit's main arguments for reductionism are successful, and ask whether his conclusions about egoism follow.

Reductionism and non-reductionism begin as accounts of when two person-stages are stages of the same person; that is, both offer accounts of the relation of co-personality. Though Parfit characterizes the dispute between reductionism and non-reductionism in different ways, an important difference, for our purposes, is that whereas reductionists think that co-personality just consists in the holding of certain other familiar physical or psychological relations among these person-stages, non-reductionists claim that co-personality does not just consist in other, more basic relations but must remain a 'simple and unanalysable', 'further' fact (pp. 210, 240). Parfit defends a version of *psychological* reductionism, according to which personal identity just consists in psychological continuity and/or connectedness, with the right kind of cause, or 'relation R', as he calls it (p. 215). Of these two relations, psychological connectedness is explanatorily prior (cf. pp. 204–9). Person-stages p1 and p2 at times t1 and t2 are psychologically *connected* just in case the psychological states of the later stage are causally dependent, in the right way, upon the psychological states of the earlier stage. For example, different stages in a life will be connected in so far as actions at one time fulfil prior intentions, goals held at one time reflect prior deliberations, memories at one time depend upon past experiences, and other beliefs and desires are maintained from one stage to the other. Well-connected selves are likely to be similar psychologically. But psychological similar-

ity and connectedness can diminish over time. Thus, p1 can be psychologically connected to p2, and p2 can be psychologically connected to p3, but p3 need not be, and is unlikely to be, as closely connected with p1 as p2 is. A series of person-stages p1 through pn is psychologically *continuous* just in case the members of every pair of temporally contiguous stages in this series are psychologically connected. It follows that p1 and pn can be psychologically continuous even if they are not psychologically well connected or connected at all.

Parfit argues that what matters in our continued existence is not personal identity itself, but relation R (this is supposed to be demonstrated by fission cases, of the sort I discuss in sections 14–16 below). Because R includes psychological connectedness as well as continuity, and the different parts of a life can be more or less closely connected, Parfit thinks that psychological reductionism undermines the temporal neutrality of rational egoism and makes it rational for us to be less concerned about our distant future than about our near future.²⁸

10 Metaphysical Depth

Parfit claims that co-personality is metaphysically 'less deep' according to reductionism than it is according to non-reductionism. As a result, he thinks, once we accept reductionism, we should reject egoism's temporal neutrality. The idea seems to be that if identity over time is less deep, intrapersonal compensation is compromised (pp. 309–12).

We can concede that co-personality is less metaphysically deep according to reductionism, in so far as reductionism claims that co-personality can be analysed in familiar and unproblematic terms. But there is no general reason to suppose that this kind of metaphysical depth affects the justification of concern. This is because our concern about some entity or property may attach to its *functional role*, rather than to its metaphysical or compositional *analysans*. For instance, it seems that a materialist should be no less concerned to prevent pain (her own or that of another) than a dualist,²⁹ and, *ceteris paribus*, I should care just as much for mint chocolate-chip ice-cream upon learning its chemical composition. Similarly, it is not clear why becoming convinced that a future self's being me consists in its being part of a (non-branching) R-related series of selves or stages should make me any less concerned about that self. The relevant claim in the egoist rationale is that persons are metaphysically distinct and temporally extended; the rationale does not seem to require any particular metaphysical account of this fact.

Sometimes our beliefs about the metaphysical *analysans* of some entity or property will, or should, affect its role in our conceptual network

in certain ways. For instance, we may learn that some gustatory properties of mint chocolate-chip ice-cream supervene on its saturated-fat content, and this may affect its desirability. But then we need an account of how the particular *analysans* to which the psychological reductionist appeals justify a change in our attitudes. Metaphysical *depth* does not establish this.

There are other ways in which Parfit does, or might, think that reductionism undermines egoism. Let us look at these.

11 A Discount Rate for Concern

Parfit's principal argument that reductionism undermines egoism's temporal neutrality applies to the normal, non-branching cases where relation R and personal identity both obtain. Relation R is what matters. According to Parfit, relation R consists of psychological continuity and/or connectedness. But psychological connectedness is a matter of degree, and, in particular, can and does diminish over time. Because relation R is what matters and it consists of psychological connectedness as well as continuity, Parfit claims, it is not irrational for me, all else being equal, to care more about my near future than about my more distant future. In fact, he thinks that a kind of discount rate that proportions concern to psychological connectedness is rationally acceptable.

My concern for my future may correspond to the degree of connectedness between me now and myself in the future. Connectedness is one of the two relations that give me reasons to be specially concerned about my own future. It can be rational to care less, when one of the grounds for caring will hold to a lesser degree. Since connectedness is nearly always weaker over long periods, I can rationally care less about my further future. (p. 313)

As Parfit notes, this is a discount rate with respect to connectedness, but not with respect to time itself (p. 314). His discount rate should, therefore, be distinguished from the discount rate with respect to time that C. I. Lewis calls 'fractional prudence'.³⁰ Egoism is neutral with respect to time itself, and so must deny fractional prudence. But egoism's temporal neutrality is also inconsistent with Parfit's discount rate, because temporal neutrality requires a kind of equal concern for different parts of one's life. The magnitude of a good or a harm should affect its rational significance. But temporal neutrality implies that the temporal location of a good or a harm within a life should be of no rational significance. If so,

then, all else being equal, an agent should be equally concerned about goods and harms at any point in his life. In particular, if near and more distant future selves are both stages in his life, then, other things being equal, an agent should have equal concern for each, even if the nearer future self is more closely connected with his present self.

Indeed, Parfit's claim about the discount rate seems too modest. He insists only that this discount rate of concern for one's future is not irrational; he does not claim that it is rationally required. Though the friend of temporal neutrality must deny the more modest claim as well, the reductionist argument, if successful, surely supports the stronger claim that a discount rate of concern is rationally appropriate where the relations that matter hold to a reduced degree. This is because concern should track and be proportional to the relations that matter.

Such a discount rate is incompatible with egoism's temporal neutrality. But reductionism does not imply a discount rate, and we should not accept one. There are two main grounds for rejecting a discount rate.

12 Two Forms of Psychological Reductionism

Parfit's argument requires that what matters is psychological connectedness as well as continuity (C&C), for it is only connectedness that will normally diminish over time; my distant and near future selves are both continuous with my present self. Now one form of reductionism would claim that relation R consists *only* in psychological continuity (C). If it is only C that matters, then I have no less reason to be concerned about my distant future than about my near future, because each is equally continuous with my present and past selves, even if each is not equally well connected with my present self.³¹ Hence, it is not reductionism *per se*, but only Parfit's version of reductionism in terms of C&C, that threatens egoism's temporal neutrality.

Throughout much of Parfit's discussion, he is agnostic about whether the relation that matters is C or C&C (pp. 207, 262, 271, 279, 283). His only explicit argument for C&C, as against C, is that we are averse to many possible losses in psychological connectedness; thus, connectedness 'matters to us' (pp. 301–2). But our aversion to losses in connectedness does not require understanding reductionism in terms of C&C.

The clearest cases of aversion to substantial psychological change involve cases that I regard as *corruption*, where I become a less attractive person. But a natural explanation of this aversion is that it is to the *disvalue* of the psychological profile I would be acquiring, and not to loss

of connectedness *per se*. This alternative explanation of my aversion is reinforced by the fact that I presumably would not be averse to loss of psychological connectedness involved in a psychological change that I have reason to regard as an *improvement*, provided that I am responsible for the change in ways that establish psychological continuity.

Indeed, because Parfit believes that what matters is C&C, he faces a problem explaining why we have reason to improve ourselves in ways that involve significant psychological changes. If what matters is C&C, then it becomes more difficult to explain how a person has prudential or agent-biased reason, as he surely seems to, to undergo a process of 'improvement' – even though it is agreed that the person emerging from the process is better off than the person beginning the process.³² Indeed, if C&C is what matters, the more significant the improvement, the less prudential reason there is to undertake it.

At one point, Parfit suggests that my reasons to be concerned about the person who emerges when I undergo a psychological change depend upon both (1) the degree of connectedness between that person and myself prior to the change and (2) the value of the new psychological profile (p. 299). But (2) is something the egoist can happily admit – in so far as an agent is concerned to promote her own interests, she must be concerned with changes that are better for her – and requires no special claims about personal identity. And (1), I suggested, seems implausible, even when conjoined with (2), and may gain specious plausibility from its association with (2).

Failure to improve oneself is a form of imprudence. So Parfit must think that this phenomenon that we misleadingly describe as a form of imprudence is properly criticized as immoral, rather than irrational (pp. 318–20). Though the terms of criticism change, Parfit may think that the assumption that it is C&C that matters involves no net change in our evaluations. But moral criticism and rational criticism may play different roles in our overall evaluation of behaviour. Moreover, Parfit's view seems likely to affect our moral attitudes towards this phenomenon. For he must think that his version of reductionism makes interference with what we misleadingly call failure to improve oneself justifiable in terms of the harm principle, rather than as paternalism. But if, as Mill claims, interference under the harm principle is more easily justified than paternalistic interference, then Parfit must conclude that his version of reductionism makes interference with such conduct easier to justify.

These changes seem inadequately motivated. Self-improvement seems to be a paradigmatic form of agent-biased rationality. In so far as Parfit's view implies that failure to undertake such changes is not subject to such criticism, this is grounds for scepticism that C&C is what matters.

To undermine egoism's temporal neutrality, therefore, Parfit must not only defend psychological reductionism, but must provide a better argument for believing that it is C&C, rather than C, that matters.

13 Equal Concern

Moreover, Parfit's own form of reductionism fails to support a discount rate. The possibility that is supposed to undermine rational egoism is the case in which my distant future self, though continuous with my present self, is less well connected with my present self than is my near future self. This possibility does not threaten temporal neutrality even if, as Parfit assumes, connectedness as well as continuity matters. This possibility demonstrates only a fact about the relation *among the parts of my life*, not a fact about the relation *between these parts and me*.

Parfit claims to be discussing what people have reason to do, but his argument is plausible only if it is person-slices or segments, rather than persons, who are agents. Suppose I consist of a series of R-related person-stages p_1 through p_n at times t_1 through t_n (figure 6.2). The time is t_1 . Though p_5 and p_{30} are equally continuous with my present self, p_5 is better connected with my present self than p_{30} is. We can see how there is less of what matters between p_1 and p_{30} than there is between p_1 and, say, p_5 . This explains why p_1 or *some person-segment ending in p_1* should be less concerned about the welfare of p_{30} than *it* should be about the welfare of p_5 . But P is not identical with p_1 or any person-segment ending in p_1 , and the welfare of p_{30} is every bit as much part of P 's welfare as is the welfare of p_5 . So this fact about the degree of connectedness between p_1 and p_{30} does not affect P 's reasons for action. And, as I have argued (section 8), it is persons, and not person-slices or segments, who are agents. Because all parts of a person's life are equally parts of her life, even if they are R-related to each other in varying degrees, benefits and harms are equally benefits and harms to her, regardless of their temporal location in the R-related series that is her life.

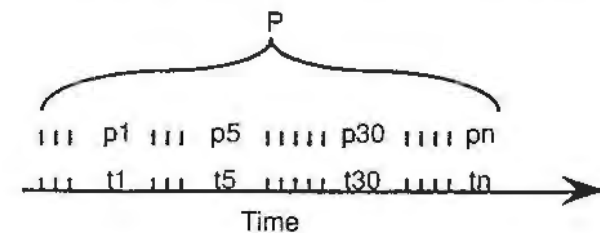


Figure 6.2 A person and her parts

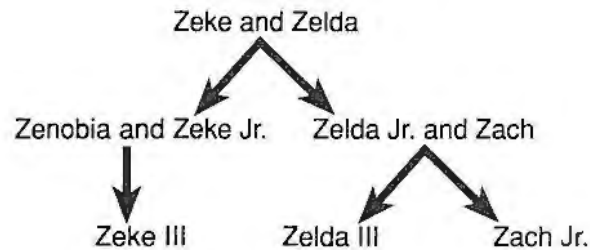


Figure 6.3 Relatives

But then diminished connectedness over time does not threaten temporal neutrality.

At one point, Parfit considers and rejects the sufficiency of an egoist appeal to this claim that all parts of a person's life are equally parts of that life (pp. 315–16). He appeals to an analogy with relatives. He claims that although all members of an extended family are equally relatives, this does not justify equal concern among them; for instance, it would not give my cousin as strong a claim to my estate as my children (p. 316). Is this a good analogy? Consider the following case (figure 6.3). Grandpa Zeke and Grandma Zelda have two children, Zeke Jr. and Zelda Jr. With his wife, Zenobia, Zeke Jr. has one child, named Zeke III; with her husband, Zach, Zelda Jr. has two children, named Zach Jr. and Zelda III. Parfit wants to say that whereas it is true that these people are all equally relatives, it does not follow that, say, Zeke Jr. should be as concerned about Zelda III (his niece) as he should be about Zeke III (his son); for instance, Zelda III does not have the same sort of claim on Zeke Jr.'s estate as Zeke III does. But to focus on the concerns of Zeke Jr. or the disposition of his estate is precisely the interpersonal (intrafamily) analogue of taking a person-stage, rather than the person, as the agent in the intrapersonal case, and I have argued against doing this. The intrafamily analogue of the person (in the intrapersonal case) is the family itself. Thus, the correct intrafamily analogue would focus on the distribution of some asset that belongs to the entire family. Here neutrality seems appropriate in light of the fact that all are equally parts of the family, even if some are more closely related than others. Thus, if Zelda Jr. is acting as the trustee of the family corporation, then, other things being equal, she ought to allocate as large a share of dividends to Zeke III (her nephew) as she does to Zach Jr. and Zelda III (her own children), even though Zach Jr. and Zelda III are more closely related to each other and to her than any of them is to Zeke III.

As long as we focus on the attitude appropriate for a whole to take towards any one of its parts, the relations that its parts bear to each other

are, *ceteris paribus*, irrelevant.³³ And as long as persons are the relevant whole, this vindicates egoism's temporal neutrality.

14 Egoism and Fission

Parfit focuses on how reductionism might undermine egoism's temporal neutrality. But one might think that reductionism undermines egoism's agent bias. Fission cases provide an independent, even if less general, argument against rational egoism. As I have explained it, egoism relies on the compensation principle. Automatic compensation seems to require that the *very same person* who is benefactor be beneficiary. This condition may not seem to be satisfied in fission cases, even though the products of fission are R-related to the subject of fission. And, as Parfit effectively argues, it seems rational for the subject of fission to be concerned about all fission products that are R-related to her (ch. 12). If so, fission may seem to undermine egoism's agent bias.

Consider the following case. Tom, Zeke and Zach are identical triplets and become involved in a serious car accident. Zeke and Zach are brain-dead; Tom is not, but his body is hopelessly mangled. Assume that it is possible to transplant Tom's brain into Zeke's body and that this preserves Tom's psychological continuity. If we do this (case 1), we regard Tom as the surviving recipient and Zeke as the dead donor (Zach is simply dead). Assume that half the brain is sufficient to sustain psychological continuity. If half of Tom's brain is seriously damaged and we transplant the healthy half into Zeke's body (case 2), Tom again survives. If, however, Tom's entire brain is healthy and we transplant half of it into Zeke's body and half into Zach's (case 3), then we have a case of fission. Call the recuperating patient in Zeke's body Dick and the one in Zach's body Harry. There is just as much psychological continuity between Tom and Dick and between Tom and Harry as there was between Tom and the recuperating patient (i.e. Tom) in cases 1 and 2. Sam and Fred are series of R-related person-stages, each of which begins with Tom's stages. The important features of fission can be represented as in figure 6.4.

As Parfit notes (p. 256), there seem to be four ways to describe what happens to Tom and his relationship to Dick and Harry.

- 1 Tom does not survive fission; in particular, he does not survive as Dick or as Harry.
- 2 Tom survives as Dick, rather than Harry.
- 3 Tom survives as Harry, rather than Dick.
- 4 Tom survives as Dick and as Harry.

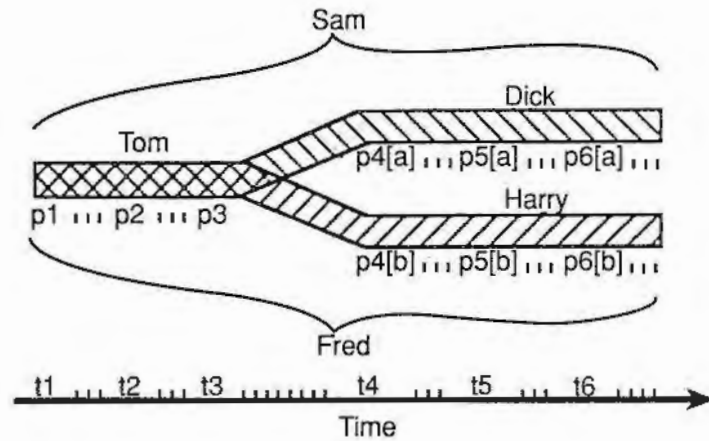


Figure 6.4 Fission

Each answer is initially hard to believe.

Against (1), we might note that there is just as much psychological continuity between Tom and Dick and between Tom and Harry as there was between Tom and the recuperating patient (i.e. Tom) in cases 1 and 2. If Tom survives in cases 1 and 2, how can he fail to survive fission? Surely, he has the same reasons to be concerned about Dick and Harry in case 3 as he did to be concerned about himself in cases 1 and 2. As Parfit says (p. 256), how can a double success be counted a failure?

But neither (2) nor (3) seems plausible. Dick and Harry have exactly equal claims to being Tom. It is true that one, but not the other, could none the less be Tom. But in virtue of what facts would one of them, rather than the other, be Tom? Dick and Harry are precisely symmetrically placed in their physical and psychological relations to Tom. If we have rejected non-reductionism, their claims to be identical to Tom must stand or fall together.

But they cannot both be Tom; (4) must be false. Identity is a transitive relation. And it seems clear that Dick is not the same person as Harry; they wake up in different hospital beds, have distinct streams of consciousness, and go on to lead different lives. But if Dick is not identical to Harry, then Tom cannot be identical with Dick and with Harry.

The best response is to accept (1) and claim that Tom does not survive fission. The transitivity of identity requires that any analysis of the relation of identity must be one-one, rather than one-many. But psychological continuity is a one-many relation. So whereas fission preserves psychological continuity, it cannot preserve identity. But because fission seems to preserve what justifies concern, we should conclude that what

principally matters as far as the rationality of concern goes is psychological continuity, rather than personal identity *per se*.³⁴

However, this response raises a problem for egoism. Because Tom survives transplants 1 and 2, he has good egoist reason to be concerned for the recuperating patient in both cases. Because Tom does not survive transplant 3, it looks as if he can have no egoist reason to be concerned about Dick or Harry; it seems he could not be compensated for any sacrifice he might make on behalf of Dick or Harry. But he bears the same (intrinsic) relationships to Dick and Harry in case 3 as he does to his recuperating self in cases 1 and 2. Fission involves *interpersonal psychological continuity*. It seems that Tom should be just (or nearly) as concerned for his fission products as for himself. If what principally matters is psychological continuity, not personal identity *per se*, then fission seems to demonstrate that an agent can have non-derivative reason to benefit others. In this way, fission might seem to undermine agent bias.

15 Egoism with Intracontinuant Compensation

If the egoist accepts reductionism, she may want to claim that maximal series of R-related person-stages – call them *continuants* – are the bearers of reasons for action. Indeed, reductionism would seem to imply that continuants are the entities with the most natural borders and, hence, are the most suitable agents (cf. section 8). The egoist could then reformulate the compensation principle in terms of continuants, rather than persons (though psychological continuants will normally be single persons).³⁵ According to this proposal, it is the two continuants Sam and Fred, rather than Tom, who have reasons for action. Because they each survive fission, each can be compensated for sacrifices that Tom might make; intracontinuant compensation will be automatic. Sam and Fred will have some common reasons and some competing reasons corresponding to their common and competing interests. We could even construct a notion of Tom's reasons out of the reasons common to Sam and Fred and an equitable accommodation between their competing reasons. For instance, we might claim that Tom has reason to take care of his health and to divide his estate equitably between Dick and Harry.

This strategy does introduce a kind of ambiguity about the subject of practical deliberation at t3, somewhat akin to the sort of ambiguities generated by the proposal that it is person-segments, rather than persons, that are the bearers of reasons for action (see section 8). But that proposal generated infinitely many possible subjects of practical deliberation in every case; the present strategy generates ambiguities only in

fission cases and, even here, only as many alternatives as there are fission products.

Some of these claims may initially sound a little strange. But in many ways fission is strange, and I cannot see anything obviously implausible in this egoist proposal.

16 Egoism with Interpersonal Compensation

Alternatively, we might retain the original formulation of the compensation principle in terms of persons, rather than continuants, and reconcile the rationality of Tom's concern with Dick and Harry and the compensation principle by claiming that Dick's good and Harry's good are *part* of Tom's good.

I think it makes sense to claim that there can be people A and B who are related such that B's good is a part or component of A's good and that, in such cases, A benefits directly from B's benefit. This sort of relationship is most commonly thought to obtain where A and B are intimates – for instance, spouses or very close friends. Where A and B are intimates, the ground for regarding B's good as part of A's good (and vice versa) seems to be that A and B interact with each other on a regular basis and help shape each other's mental life. To a large extent, the experiences, beliefs, desires, ideals and actions of each depend upon those of the other. Of course, these are the sorts of conditions of psychological continuity and connectedness that are maximally realized in the *intrapersonal* case. Here they are realized to a very large extent in the *interpersonal* case, and this is what grounds our claim that B's interests extend A's and that B's good is part of A's good (and vice versa). We might say that B stands to A as 'another-self'.³⁶ Now fission represents the limiting interpersonal case; Dick and Harry share Tom's goals and other psychological traits to the maximal extent possible, and they acquired these traits by a direct causal process from Tom. Indeed, fission cases seem to present the clearest case for claiming that one person's good is part of another's.

Unlike intimates who have ongoing interactions, Tom does not have an ongoing relationship with either Dick or Harry; indeed, he does not temporally overlap with either of them. Tom passes on his traits to Dick and Harry, and ceases to exist in the process. In this way, the relationship between Tom and Dick and Harry is highly *asymmetrical*; Dick and Harry owe their existence and nature to Tom, but not vice versa. For this reason, the parent-child relationship may present a better interpersonal model for the fission case than the relationship between spouses or friends. In many parent-child relationships the child's physical and psy-

chological nature are to a significant degree a causal product of the parent's physical and psychological nature and activities, but not vice versa. This is how we can explain the common views that the parent's interests are extended to the child's welfare, that the child's welfare is part of the parent's good, and that the parent's interests can extend beyond her own existence. The continuity between Tom and both Dick and Harry is even greater than that between parent and child; this is our ground for claiming that Dick and Harry extend Tom's interests and for claiming that Dick's good and Harry's good are part of Tom's.

Of course, the asymmetrical character of the psychological relationship is clearest in the normal, intrapersonal case. In the normal, non-branching case, relation R both extends the agent's interests and extends the agent. Tom is a person who consists of a non-branching series of psychologically continuous person stages p1 through p3. This psychological continuity from p2 to p3 extends Tom's interests in the sense that p3 inherits, carries on, and carries out p2's projects and plans. In this non-branching case, it also extends Tom's life. In the fission case, however, relation R does not literally extend Tom's life, because neither Dick nor Harry is Tom. But, by virtue of being fully psychologically continuous with Tom, Dick and Harry will each inherit, carry on, and carry out Tom's projects and plans (though presumably in somewhat different directions over time). This seems to be a good ground for claiming that Dick and Harry extend Tom's interests, in the very same way that his own future self would normally extend his interests, even if they do not literally extend his life.

If so, interpersonal psychological continuity extends the agent's interests in much the way that intrapersonal psychological continuity does. An agent's future welfare is a constituent of her overall good. As such, she has reason to be concerned about her future for its own sake, as part of her overall good. In a similar way, if another's good is a constituent of the agent's overall good, then she has reason to be concerned about the other for his own sake, as part of her overall good. If so, the egoist can recognize derivative but non-instrumental reason to be concerned about others.

Nor is the scope of this egoist justification of other-regarding concern limited to one's fission products or intimates. Dick and Harry are, by hypothesis, maximally continuous with Tom. What makes fission a case of interpersonal psychological continuity is that continuity takes a one-many form. But we have seen that there is interpersonal psychological continuity in other cases – for instance, between intimates. If so, what distinguishes between future stages of myself and future stages of my intimates? The answer, according to psychological reductionism, must be that there is more continuity between myself now and myself in the

future than between myself now and my intimate in the future. According to psychological reductionism, what distinguishes intrapersonal continuity and interpersonal continuity (in non-branching contexts) is the degree of continuity. There are more numerous and more direct psychological connections – between actions and intentions and among beliefs, desires and values – in the intrapersonal case. And where the connections between links in a chain are all weaker, continuity between any points in the chain will also be weaker. If so, we can see that I am more weakly continuous with my intimates than I am with myself. We can also see how I might be continuous with others, besides my intimates, even if more weakly so. I interact directly with others, such as colleagues and neighbours, and this interaction shapes my mental life in certain ways, even if the interaction in such cases is less regular than is my interaction with intimates, and even if the effect of such interactions on my mental life is less profound than is the effect produced by interaction with my intimates. Moreover, I interact with a much larger network of people indirectly, when the psychological influence between me and them is mediated by other people and complex social institutions, though the continuity thus established is, as a result, weaker. Indeed, the nature of my relationships to others and of the bonds between us are a function of the degree of interpersonal psychological continuity between us. If an agent's interests are extended in so far she is psychologically continuous with others, then the egoist can justify derivative but non-instrumental concern for others that is proportional to the amount of psychological continuity that exists between the agent and others.³⁷ If so, the egoist can justify a familiar kind of interpersonal discount rate. The central claims of self-referential altruism about other-regarding conduct can stand, not as axioms of practical rationality, but as theorems derived from an agent-biased theory.³⁸

17 Conclusion

Rational egoism has a rationale for its asymmetrical treatment of intertemporal and interpersonal distribution that relies on the separateness of persons. The separateness of persons suggests the principle that it is unreasonable for agents to make uncompensated sacrifices. If we apply the compensation principle in both intrapersonal and interpersonal contexts, we explain temporal neutrality and agent bias; intrapersonal compensation is automatic, whereas interpersonal compensation is not. But then there is nothing arbitrary about egoism's hybrid structure; at a deeper level, it treats intertemporal and interpersonal distribution the same. I have tried to defend this rationale against

two main challenges that Parfit's discussion naturally suggests. One might defend the fully biased presentism by arguing that it is person-slices or segments, rather than persons, that are agents and subjects of compensation. But it is persons that engage in practical deliberation and possess reasons for action; if so, temporal neutrality assures intrapersonal compensation. Though the egoist rationale does appeal to the separateness of persons, it is not undermined by reductionist accounts of personal identity. The reductionist should accept temporal neutrality; all else being equal, agents should care equally about different parts of their lives, even if these parts are not equally related to each other. Nor does the reductionist account of fission undermine agent bias. If we continue to think of agents as persons, concern for those to whom one is R-related can be reconciled with agent bias by recognizing how interpersonal psychological continuity makes interpersonal compensation possible. Alternatively, if we think of agents as continuants, rather than persons, then concern for others to whom one is R-related can be reconciled with agent bias by recognizing how interpersonal psychological continuity makes intracontinuant compensation possible. If so, rational egoism's commitment to the separateness of persons explains its asymmetrical treatment of interpersonal and intertemporal distribution in a way that is metaphysically robust.³⁹

Notes

- 1 Henry Sidgwick, *The Methods of Ethics*, 7th edn. University of Chicago Press, Chicago, 1907; hereafter *ME*.
- 2 Thomas Nagel, *The Possibility of Altruism* (Princeton University Press, Princeton, 1970), esp. pp. 16, 19, 99–100.
- 3 Parfit is attracted to what he calls a *Critical* version of P, CP, which tells each agent to aim at what will best achieve his present rational aims (pp. 94, 118–19). Rational aims exclude any intrinsically irrational desires, and include any desires that are rationally required. Because any aims might, in principle, turn out to be intrinsically irrational or rationally required, Parfit concludes that any possible theory of rationality can be represented as a version of CP (pp. 193–4). But these claims about CP are misleading. First, CP is not best represented as a version of P. Aims that are rationally required may concern things that are not the object of my actual or counterfactual desires; though I should desire them, I may not, not even under ideal epistemic conditions. But in so far as 'my rational aims' concern things that may satisfy none of my desires, CP should simply define rational action in terms of achieving those things now, not in terms of my present aims or desires. In this way, CP is more akin to presentism than to P. Nor is it helpful to think of CP in terms so broad that any theory counts as a possible version of CP.

- as long as CP is a rival to other theories and is fully relative (biased). For these purposes, CP is best understood as a version of presentism with an objective theory of value.
- 4 See Thomas Nagel, *The View from Nowhere* (Oxford University Press, New York, 1986), pp. 152–3.
 - 5 See C. D. Broad, 'Self and Others', repr. in *Broad's Critical Essays in Moral Philosophy*, ed. D. Cheney (George Allen & Unwin, London, 1971), pp. 262–82.
 - 6 On an instrumental view, as Hume famously observes, 'Tis not contrary to reason to prefer the destruction of the whole world to the scratching of my finger. 'Tis not contrary to reason for me to chuse my total ruin, to prevent the least uneasiness of an *Indian* or person wholly unknown to me. 'Tis as little contrary to reason to prefer even my own acknowledg'd lesser good to my greater, and to have a more ardent affection for the former than the latter' (David Hume, *A Treatise of Human Nature*, ed. P. Niddich (Clarendon Press, Oxford, 1978), II.iii.3, p. 416).
 - 7 An instrumental conception of P might also be thought to forge a link between normative or justifying reasons and explanatory reasons. We explain an agent's behaviour when we 'rationalize' it as an attempt to satisfy certain desires she had, given her beliefs. This may suggest that behaviour that is rationally justified is behaviour that would fulfil the agent's desires, or at least those desires she would have if she met certain epistemic conditions.
 - 8 The fact that Parfit's preferred version of P is CP, in which the agent's aims sometimes play no essential explanatory role, makes it clearer that his fully relative P can, and should, be understood as presentism (cf. n. 3 above).
 - 9 Nagel has revised his views in *View from Nowhere*, chs 8–9. Whereas *Possibility of Altruism* denied that there are agent-relative reasons, *View from Nowhere* recognizes some agent-relative reasons.
 - 10 Nagel, *Possibility of Altruism*, pp. 16, 19, 99–100.
 - 11 *Ibid.*, ch. 13, esp. pp. 134–42.
 - 12 Contrast John Perry, 'The Importance of Being Identical', in *The Identities of Persons*, ed. A. Rorty (University of California Press, Los Angeles, 1976), pp. 67–90.
 - 13 If personal desires essentially refer to the person who has them, then they appear to presuppose personal identity. If so, it may seem that personal desires cannot help constitute personal identity. The apparent circularity of treating personal desires as an ingredient of personal identity is like the circularity of treating memory as an ingredient of personal identity, and the solution is the same. We can appeal to 'quasi-memory', which is like memory except that it presupposes continuity, rather than identity, in our account of personal identity (p. 220). So, too, we can appeal to 'quasi-personal desire', which is like personal desire except that it presupposes continuity, rather than identity, in our account of personal identity.
 - 14 Cf. Jennifer Whiting, 'Friends and Future Selves', *Philosophical Review*, 95 (1986), p. 579.
 - 15 See Michael Bratman, *Intention, Plans, and Practical Reason* (Harvard University Press, Cambridge, Mass., 1987).

- 16 A similar claim linking the separateness of persons with a claim that the morality of sacrifice requires compensation is a familiar theme from recent moral philosophy. In charging that utilitarianism fails to recognize the separateness of persons, critics often complain that it allows interpersonal balancing, and so permits uncompensated sacrifices. (Cf. Nagel, *Possibility of Altruism*, pp. 138, 142; John Rawls, *A Theory of Justice* (Harvard University Press, Cambridge, Mass., 1971), pp. 23–4, 26–7, 29, 187–8, 191; and Robert Nozick, *Anarchy, State, and Utopia* (Basic Books, New York, 1974), pp. 31–4.) However, we should find the compensation principle more plausible as applied to rationality than as applied to the morality of sacrifice, to the extent that we regard the moral point of view as an impartial one and the rational point of view as the agent's personal point of view. For discussion of the compensation principle as applied to morality, see my 'The Separateness of Persons, Distributive Norms, and Moral Theory', in *Value, Welfare, and Morality*, ed. R. Frey and C. Morris (Cambridge University Press, New York, 1993), pp. 252–89.
- 17 Recognition of this fact may tend to undermine the force of some of Parfit's other criticisms of egoism, on the ground that its temporal neutrality requires us to act now on past desires that we no longer have and on future desires that we do not yet possess (cf. ch. 8, esp. sects 59–61). But to pursue these issues would require a fuller discussion of the commitments of temporal neutrality than I have space for here (I hope to pursue these issues elsewhere).
- 18 I explore egoist strategies for justifying other-regarding conduct and concern more fully in my 'Rational Egoism, Self, and Others', in *Identity, Character, and Morality*, ed. O. Flanagan and A. Rorty (MIT Press, Cambridge, Mass., 1990), pp. 339–78, and 'Self-love and Altruism', *Social Philosophy and Policy*, 14 (1997), 122–57.
- 19 See my 'Rational Egoism, Self, and Others' and 'Self-love and Altruism'.
- 20 Nagel, *Possibility of Altruism*, p. 142.
- 21 For example, Parfit raises some interesting questions about temporal neutrality (Pt 2, ch. 8), which I cannot pursue here.
- 22 Sidgwick says: 'it must surely be admissible to ask the Egoist, "Why should I sacrifice a present pleasure for a greater one in the future? Why should I concern myself about my own future feelings any more than about the feelings of other persons?"' (ME, p. 418). Parfit says: 'A Present-aim Theorist can ask, "Why should I give weight *now* to aims which are not mine *now*?"' (p. 95).
- 23 Cf. Mark Heller, 'Temporal Parts of Four-dimensional Objects', *Philosophical Studies*, 46 (1984), pp. 323–34.
- 24 For example, Williams bases his reservations about temporal neutrality on the thought that 'The correct perspective on one's life is *from now*' (Bernard Williams, 'Persons, Character, and Morality', repr. in his *Moral Luck* (Cambridge University Press, New York, 1981), p. 13).
- 25 Cf. Christine Korsgaard, 'Personal Identity and the Unity of Agency: A Kantian Response to Parfit', *Philosophy and Public Affairs*, 18 (1989), pp. 101–32.

- 26 Perhaps the difficulty arises only if we are realists about temporal parts, and perhaps the proper moral of the difficulty is that we should reject realism about temporal parts. The defence of presentism that I am considering in this section presupposes a realism about temporal parts. If we reject realism about temporal parts, this hurts presentism, not egoism.
- 27 Joseph Butler, *The Analogy of Religion*, appendix 1, repr. in *Personal Identity*, ed. J. Perry (University of California Press, Los Angeles, 1975), p. 102.
- 28 I will not discuss the 'Extreme Claim', which Butler accepts and Parfit discusses (pp. 307–12), according to which reductionism completely undermines any special concern for oneself and one's future. Because I think that egoism survives even Parfit's more moderate claim, there's not much reason to discuss the extreme claim directly.
- 29 Cf. Sydney Shoemaker, Critical Notice of Derek Parfit, *Reasons and Persons*, *Mind*, 94 (1985), p. 451 (repr. as ch. 7 below).
- 30 See C. I. Lewis, *An Analysis of Knowledge and Valuation* (Open Court, La Salle, Ill., 1946), p. 493.
- 31 I do think that continuity, like connectedness, can be a matter of degree. If, of two chains A and B, the links in A are better connected than the links in B, then the end-points in A will be more continuous than the end-points in B. However, this does not affect my present point, which assumes only that there is continuity within a chain despite a constant rate of diminished connectedness.
- 32 Compare Aristotle's discussion of the puzzle about whether to wish one's friend the good of divinity in *Nicomachean Ethics*, tr. T. Irwin (Hackett, Indianapolis, 1985), VIII 7, 1159a5–11. Aristotle claims that one who cares about the friend for the friend's own sake would not wish this good on him, because the friend would not survive the transformation. While Aristotle's claim seems plausible, C&C's corresponding claims about the (ir)rationality of self-improvement seem implausible. Aristotle's claim is plausible, because it is plausible to think that the species or genus to which an individual belongs is an essential property of that individual; whereas it seems implausible to think that an individual's current psychological profile is essential to that individual's persistence in the way that C&C seems to require.
- 33 Of course, all else is not equal if the relations among the parts of a whole are such that a given part is a less important part of the whole than others (e.g. as my fingernail is a less important part of my body than my heart). In such a case, treatment of the parts as equals may not support equal treatment of the parts. But notice that this will be a result of the relation between that part and the whole, and not simply between that part and other parts. Notice too that this is compatible with temporal neutrality.
- 34 If so, Parfit is wrong to claim, as he does (pp. 259–60, 278–9), that it is an 'empty question' which of the four answers is right, because they all describe the same outcome. I am unsure myself whether psychological continuity, rather than identity, is all that matters; if it can matter that psychological continuity takes a unique or non-branching form, then identity will have some independent value. However, it is enough to get this challenge to

- egoism off the ground that psychological continuity has significant independent value.
- 35 This suggestion is similar to Lewis's view; see David Lewis, 'Survival and Identity', in *Identities of Persons*, ed. Rorty, esp. pp. 24–9. Lewis seems to want to identify persons and continuants; however, I do not. I think that brains or brain parts and the psychological continuants they support make up one person if and only if they are *functionally integrated*.
- 36 This phrase comes from Aristotle's justification of friendship in *Nicomachean Ethics*, bk IX, 1166a30–2, 1170b6–8. For further discussion of Aristotle's justification of friendship and its relevance to egoist justifications of other-regarding concern, see T. H. Irwin, *Aristotle's First Principles* (Clarendon Press, Oxford, 1988), ch. 18.
- 37 I discuss various issues about the scope and weight of the reasons that emerge from this sort of justification of other-regarding conduct and concern in 'Self-love and Altruism'.
- 38 In fact, this egoist justification of concern for others to whom one is R-related allows an egoist justification of temporal neutrality even on the assumption (rejected in sect. 8) that person-segments, rather than persons, are agents. Person-segments have reason to care about future segments to whom they are R-related. Assume, contrary to fact, that person-segments are agents. Temporally contiguous person-segments will be strongly R-related, and so the earlier segment will have egoist reason to benefit the later segment, even if C&C matters, despite the non-identity of the person-segments. In the case of temporally distant person-segments that are psychologically continuous but not well connected, the earlier segment will have egoist reason to benefit the later segment just in case C alone matters.
- 39 *Postscript* (added August 1995). Whereas my commitment to temporal neutrality requires me to reject an intrapersonal discount rate (sects 11–13), I think I would like to accept an interpersonal discount rate, not unlike the one recognized by self-referential altruism, as at least part of the correct view about rationality and morality (sect. 16). This asymmetry need not be problematic, as I hope to explain. But it may be hard to justify a familiar interpersonal discount rate if we believe that it is continuity (C), rather than continuity and connectedness (C&C), that matters. Continuity can admit of degrees; if of two chains A and B the links in A are better connected than the links in B, then any points in A will be more continuous than any points in B. So, for example, I am better connected and more continuous with myself in the future than I am with others, even intimates. But I may be equally continuous with many others, even if I am not equally well connected with all of them. This can be illustrated in terms of friendship. Let us say that friends are friendship-connected. Suppose that X is friends only with Y, whereas Y is friends also with Z. All else being equal, X is as friendship-continuous with Z as she is with Y, though she is friendship-connected to Y but not to Z. A familiar interpersonal discount rate, of the sort recognized by self-referential altruism, would claim that, all else being equal, X has more reason to be concerned about her friend Y than about her

friend's friend Z. We can endorse this claim if both C&C matter. Apparently, we cannot endorse it if only C matters, for X's friend Y and her friend's friend Z are equally continuous with X. So whereas if C alone matters, we could justify a self-other asymmetry, we could apparently not justify a familiar interpersonal discount rate. For this, connectedness, as well as continuity, must matter. Defending an interpersonal discount rate this way may seem inconsistent with my rejection of an intrapersonal discount rate (sects 11–13). In one way, it is; in another way, not. In this essay I offered two reasons for rejecting an intrapersonal discount rate. First, I pointed out that Parfit's argument for an intrapersonal discount rate requires that it be C&C, rather than C alone, that matters, as only connectedness normally diminishes over the course of a life. But, I argued, we might well suppose that it is C, rather than C&C, that matters, in part because this seemed necessary to justify prudential reasons for self-improvement (sect. 12). Second, I argued that Parfit's argument for an intrapersonal discount rate does not work even if C&C both matter, because it confuses parts and wholes (sect. 13). But it seems that I cannot appeal to the first argument against Parfit's discount rate and appeal to the significance of C&C to defend an interpersonal discount rate. I am now inclined to think that it was mistaken and unnecessary to suppose that the first argument should stand on its own. I still think that Parfit offers no good reasons for thinking that it is C&C, rather than C, that matters, and that his argument for a discount rate has a problem justifying prudential reasons for self-improvement. But we can avoid a discount rate and problems about the prudential justification of self-improvement and defend temporal neutrality, without denying that connectedness as well as continuity matters, provided we avoid what I think is Parfit's confusion of parts and wholes. For we can give prudential reasons for undertaking psychological improvements that diminish connectedness among one's parts by appeal to the need for a person to have equal concern for all her parts, regardless of the relations among them (sect. 13). This requires me to rest my case against the intrapersonal discount rate on the second argument. It would also allow me to accept an interpersonal discount rate while rejecting an intrapersonal one.

7

Parfit on Identity

Sydney Shoemaker

Each of the four parts of Derek Parfit's impressive and important book deserves detailed examination, and nothing short of another book could give detailed examinations of them all. Here I will focus exclusively on the discussion of personal identity that constitutes Part 3 of the book. The broad outlines of this account will be familiar to those who have read Parfit's well-known 1971 paper 'Personal Identity'.¹ The two targets of that paper – the view that personal identity is perfectly determinate (that questions of personal identity always admit of a yes or no answer) and the view that 'what matters' in survival is personal identity itself – are also among the main targets here. Parfit believes that we are naturally disposed to accept a 'non-reductionist' account of personal identity. According to this account, persons are 'separately existing entities', whose existence is all-or-nothing and does not consist in the holding of certain relations between mental events and bodies, and whose identity is perfectly determinate. Unity of consciousness is explained in terms of 'ownership' of different experiences by such a separately existing entity. And it is the continued identity of the entity of this sort one is that 'matters' – this is the focus of the special concern one has for one's future existence and well-being. But while this is what we tend to believe, according to Parfit, it is not what we should believe. Parfit champions a 'reductionist' account according to which we are not such separately existing entities, and according to which personal identity consists in facts that can be described 'impersonally', more specifically in 'non-branching psychological continuity and connectedness' (I shall sometimes abbreviate this 'psychological C&C'). We have psychological continuity when a person remembers his earlier deeds or experiences, or when an intention formed at one time is fulfilled at a later time, or when

First published as a Critical Notice of *Reasons and Persons* in *Mind*, 94 (1985), pp. 443–53. The present version is slightly abridged.

Reading Parfit

Edited by
Jonathan Dancy

 **BLACKWELL**
Publishers