PHIL 131: Topics in Metaphysics  
Spring 2015; David O. Brink  
Topic: Persons & Values  
Handout #5: Reductionism, Rationality, and Prudence

In *Reasons and Persons* Parfit famously claims that accepting the truth of psychological reductionism should be normatively revisionary, modifying our views of practical reason and morality. In this respect, he seems to agree with Butler and Reid that a Lockean conception of personal identity, at least in general outline, if not in the details, has revisionary normative implications. Here, I want to focus on his claims about the bearing of reductionism on our conception of practical reason.

### STRUCTURAL CONCEPTIONS OF PRACTICAL REASON

This focus on reductionism and rationality is part of a larger discussion in *Reasons and Persons* about rationality and morality. Parfit is especially interested in three different structural conceptions of practical reason that vary in their attitudes about the distribution of goods and bads both across persons and across time. A conception is *neutral* with respect to persons or time if it assigns no intrinsic significance to how goods and bads are distributed among persons or over time. By contrast, we might say, a conception is *biased* or *relative* if it does assign intrinsic significance to issues about whom a benefit or burden falls upon or when it does. Though there could in principle be many kinds of bias, Parfit focuses on a bias toward the agent himself, in relation to others, and bias toward the present. This would seem to yield four possible conceptions of practical reason.

<table>
<thead>
<tr>
<th>Persons</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutrality</td>
<td>Neutrality/Rational Benevolence</td>
</tr>
<tr>
<td>Bias</td>
<td>Egoism/Prudence/Self-interest</td>
</tr>
</tbody>
</table>

These are *structural* conceptions of practical reason that make claims about the distribution of goods and bads across persons and times. They are agnostic about what the good consists in, as between hedonism, desire-satisfaction, perfectionism, or objective list views. Parfit focuses only on three of these four – the fully neutral Rational Benevolence (which he also calls Neutralism), the fully biased or relative Presentism (which he calls the Present-aim Theory), and the hybrid Egoism or Prudence (which he calls the Self-interest Theory). In doing so, he ignores the other hybrid, Benevolence of the Present Moment, perhaps because he thinks it’s a possibility no one has endorsed or wants to endorse.

Part II of *Reasons and Persons* contains several ingenious arguments against egoism. These arguments repay careful study. But as most of them do not presuppose any particular view of personal identity, we cannot afford to give them the attention that they otherwise would deserve. For the most part, with a couple of exceptions, I will confine myself to summarizing Parfit’s arguments.

### AGENT BIAS

One argument focuses on egoism's agent-bias. Egoism implies that it might be irrational to act on other-regarding moral sentiments if this is contrary to the agent’s self-interest, whereas the
present-aim theory need not condemn such action if it appeals to the desires the agent has at the time of acting (§50). Parfit thinks that this is problematic for egoism and a feather in the cap of the present-aim theory.

But the present-aim theory can’t explain why people who lack tuistic desires are irrational or those who happen to have them should retain them. By contrast, we could try to answer both questions by appeal to the agent’s enlightened self-interest. It is, of course, controversial how successful the egoist answer is.

TEMPORAL NEUTRALITY

Other arguments in Part II of Reasons and Persons attack egoism’s temporal neutrality. Most of these are fascinating arguments. Though I do not regard any of them as decisive, the best responses to them defy easy summary.\(^1\) Since most of these arguments do not presuppose any particular view of personal identity, their detailed reconstruction and assessment are beyond the scope of our focus in this course.

PAST DESIRES

One worry is that if we attach normative significance to satisfying one’s desires, as most conceptions of practical reason do in one way or another, then temporal neutrality requires us to assign significance to the satisfaction of past desires, even ones we no longer hold but now repudiate (§61), which may seem like a problematic commitment.

But this is a problem only if we assume that what has normative significance is the satisfaction, at any time, of any desire one holds, at any time. Why should we assume that? It seems much more plausible to construe the significance of desire satisfaction as accruing to the satisfaction of desires that the agent holds at the time of satisfaction. But the temporally neutral version of this claim does not seem especially problematic.

INTRAPERSONAL CONFLICTS OF IDEALS

A related worry is that temporal neutrality requires acting out of deference to desires, projects, and ideals that one does not now hold but that will hold in the future (§60). Indeed, temporal neutrality might require me to act out of deference to ideals that I do not now hold and in fact now repudiate. The stock example is that of the young radical. Temporal neutrality seems to counsel that she moderate pursuit of her radical ideals if she will have staid bourgeois ideals of security, stability, and material comfort later in life. This may seem to require the impossible or at least require her to sacrifice her integrity and authenticity.

Since Parfit thinks that the worry is even more acute if we are psychological reductionists about personal identity, we might do best to postpone our discussion of this criticism of temporal neutrality until we consider Parfit’s specifically reductionist claims about rationality (see below).

THE SYMMETRY ARGUMENT

Another interesting worry about temporal neutrality comes out in the context of a famous Epicurean argument for why we should not fear death. The Epicureans saw the main aim of philosophy as confronting and, if possible, removing the fear of death, which, as hedonists, they regarded as bad insofar as it causes anxiety. They thought that fear of death was predicated largely on fear of retribution from anthropomorphic gods. They offered many different sorts of arguments for why we should not fear death. They argued that if the gods do exist we have reason to think that they do not interfere in human affairs and that even if they do exist and intervene in human affairs we are invulnerable to harm after death. Some of these arguments assume death brings

---

nonexistence. Others do not. The argument that bears on temporal neutrality purports to show that we have no reason to fear death even if — indeed, because — it implies our nonexistence. In *De Rerum Natura* Lucretius gives expression to temporal neutrality in appealing to a parallel between our prenatal and postmortem nonexistence to counteract our fear of death.

From all this it follows that death is nothing to us and no concern of ours, since our tenure of mind is mortal. In days of old, we felt no disquiet when the hosts of Carthage poured into battle on every side — when the whole earth, dizzied by the convulsive shock of war, reeled sickeningly under the high ethereal vault, and between realm and realm the empire of mankind by land and sea trembled in the balance. So, when we shall be no more — when the union of body and spirit that engenders us has been disrupted — to us, who shall then be nothing, nothing by any hazard will happen any more at all. Nothing will have the power to stir our sense, not though earth be fused with sea and sea with sky [III 830-51].

Later, he invokes the same symmetry between postmortem and prenatal nonexistence.

Look back to see how the immense expanse of past time, before we were born, has been nothing to us. Nature shows us that it is the mirror-image of the time that is to come after we are dead. Is there anything there terrifying, does anything there seem gloomy? Is it not more peaceful than any sleep [III 972-77]?

The Symmetry Argument is wonderful. Here is its structure.

1. Death brings nonexistence.
2. Postmortem nonexistence is no different than prenatal nonexistence.
3. We do not regret our prenatal nonexistence.
4. Hence, we should not regret our death.

The Epicureans notice an asymmetry in our attitudes toward past and future nonexistence. They reject this asymmetry as irrational and propose to make our attitudes toward death consistent with our attitudes toward prenatal nonexistence. In so doing, they embrace temporal neutrality.

But symmetry is a two-edged sword. The parity of prenatal and postmortem nonexistence could be exploited to expand, as well contract, regret. Consider this second appeal to symmetry.

1. Death brings nonexistence.
2. Postmortem nonexistence is no different than prenatal nonexistence.
3. We do regret our death.
4. Hence, we should regret our prenatal nonexistence.

This second appeal to symmetry may seem more compelling if we have no independent explanation of why death is not bad. Of course, the Epicureans also appeal to an Existence Requirement – one cannot be harmed if one does not exist – to explain why nonexistence is not to be feared (III 860-70). But the Existence Requirement does not explain why death is not bad. Even if one cannot be harmed after death, one can be harmed by death, because death deprives the person whom it befalls of the goods she would have enjoyed had she continued to exist and led a life worth living. If this is what is bad about death, then symmetry suggests that we do have reason to regret our prenatal nonexistence. Had we existed earlier (and lived to the same date as we actually do), we would have enjoyed more goods than we will in fact. Either form of nonexistence deprives us of possible goods and so is a legitimate source of regret. Of course, to say that death or prenatal nonexistence is an appropriate object of regret is not to endorse preoccupation with it.
Some may regard either symmetry argument as a reductio of temporal neutrality. Parfit seems sympathetic with this response (§67). However, the symmetry that the Epicureans appeal to is surprisingly robust, whether we use it to contract or expand regret.2

OUR PREFERENCE FOR PAST PAIN

Still another interesting worry about temporal neutrality is that our preferences, even our enlightened or considered preferences, don’t seem to track this constraint. Most of us would prefer learning that our suffering is past, even if this suffering is greater than would be an alternative future suffering. Parfit illustrates this claim with his ingenious case of My Past and Future Operations (§64). Imagine that there is a painful operation that requires the patient’s cooperation and, hence, can only be performed without the use of anesthetic. But doctors can and do induce (selective) amnesia after the operation to block memories of these painful experiences, which are themselves painful. I knew I was scheduled for this procedure. I wake up in my hospital bed and ask my nurse whether I have had the operation yet. He knows that I am one of two patients, but doesn’t know which. Either I am patient A, who had the longest operation on record yesterday (10 hours), or I am patient B, who is due for a short operation (one hour) later today. While I wait for him to check the records, I find that I have the strong preference and hope that I am patient A, even though A’s suffering was greater than B’s will be. Temporal neutrality would seem to imply that this preference is irrational. But that might not seem right. More generally, it might seem that we prefer to minimize future suffering, even if that is not a way to minimize total suffering.

EGOISM’S HYBRID STRUCTURE

A final worry about egoism, which is worth some discussion, appeals not to its agent-bias or to its temporal neutrality but to its combination of the two. This is a worry about the hybrid structure of egoism. Time and person are parallel distributional dimensions; we need to decide where to locate goods and evils in time and among persons. Once we adopt this perspective, prudence may seem like an unstable hybrid. It says 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. In The Methods of Ethics Sidgwick considers this issue in the context of his discussion of 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?’ [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 the same. Parfit pushes this same worry about the hybrid structure of prudence, or the self-interest theory (S), as he calls it, in Part II of Reasons and Persons.

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 [140].

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 the view Thomas Nagel adopts in The Possibility of Altruism. His primary aim is to argue against egoism's agent-bias and in favor 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 reason 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 (16, 19, 99-100).

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 (RP 137-44).

Whereas Parfit thinks that one cannot defend the hybrid character of prudence, Sidgwick thinks that this challenge to prudence is unanswerable only if we accept Humean skepticism about personal identity over time (Methods 418-19). Sidgwick thinks that prudence is defensible provided we recognize the separateness of persons.

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 [498].

This appeal to the separateness of persons suggests a rationale for the hybrid structure of prudence. We saw that when the separateness of persons is invoked to discredit utilitarianism critics of utilitarianism appeal to the compensation principle. But the compensation principle and the metaphysical separateness of persons explain the asymmetry between intrapersonal and interpersonal distribution. We saw that 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 relativity or bias. So we have a rationale for the hybrid treatment prudence accords intertemporal and interpersonal distribution.

Or do we? Couldn’t doubts about interpersonal balancing be extended to intrapersonal balancing? If the separateness of persons defeats interpersonal balancing, why doesn’t the separateness of different periods within a person’s life defeat intrapersonal balancing? After all, me-now and me-later are distinct parts of me. But then it is hard to see how me-now is any more compensated for its sacrifices on behalf of me-later than I am compensated by my sacrifices for you. Just as doubts about interpersonal balancing lead to a distributed concern with each person, perhaps doubts about intrapersonal balancing should support a distributed concern with each part of a person’s life. There are different interpretations of what this distributed concern requires in the interpersonal context, such as equal distribution and maximin. Perhaps we need to explore comparable interpretations of distributed concern in the intrapersonal context. However, this concern about temporal neutrality is not compelling, as it stands, for several reasons.

First, we might distinguish between temporal impartiality and temporal neutrality. Consider again the interpersonal case. Here, one norm might be called the norm of impartiality; it insists that everyone be given equal concern. This norm of impartiality admits of different interpretations, including a norm of substantive equality and maximin, among others. Indeed, utilitarianism’s person neutrality is one interpretation of interpersonal impartiality. Similarly, we might identify a more generic notion of intertemporal impartiality that would admit of different interpretations, including that of temporal neutrality. One way to read the separateness argument, then, is to see it mandating a temporal impartiality. That would not vindicate temporal neutrality, as such, but it would require a form of impartiality that was inconsistent with the sort of temporal bias displayed in ordinary life by familiar forms of temporal discounting and displayed theoretically in the purebred presentism.

Second, this challenge to temporal neutrality requires thinking that we can and should adopt a sub-personal perspective when reckoning compensation. But there are problems with this idea. Once we go sub-personal and appeal to full relativity, there seems no reason to stop until we reach the sub-personal limit – a momentary time slice of the person. But notions of compensation have no application to momentary time slices, which do not persist long enough to act or receive the benefits of earlier actions. Moreover, many of the goods in life, especially the pursuit and achievement of worthwhile projects, seem to be realized only by temporally extended beings. But if we stop short of momentary time slices and appeal to larger sub-personal entities, call these person segments, other problems arise. One question is just where to stop. If we don’t fully relativize, why relativize partially? Moreover, if we do relativize partially, we introduce indeterminacy. This is because the careers of person segments overlap, with the result that any one point in time is part of the career of indefinitely many different segments. To decide whether compensation has occurred, we need a determinate subject. But if we appeal to person segments, we seem to lack a determinate subject.

Of course, persons are just maximal segments. They also seem to be the most salient segments. Many of the things we value and that structure our pursuits are certain sorts of lives. We aim to be certain sorts of people. Insofar as these ideals structure our beliefs, desires, and intentions, the correct perspective from which to assess success would seem to be the perspective of a whole life. Even when persons have more parochial aims and ambitions, the successful pursuit of these aims and ambitions requires interaction and cooperation among segments, much as persons must often cooperate with others to achieve individual, as well as collective, aims. They do interact and cooperate, much as distinct individuals interact and cooperate in groups, in order to plan and execute long-term projects and goals. They must interact and cooperate 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 cooperate if they are to use scarce resources to mutual advantage. Indeed, both the ease and necessity of interaction among person
segments will be greater than that among persons, because the physical constraints and the reliability of fellow cooperators 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 we should keep going until we reach an entity with the most natural borders, viz. the person.

REDUCTIONISM AND RATIONALITY

In Part III of Reasons and Persons Parfit claims that psychological reductionism provides independent arguments against egoism. His main focus here seems to be on egoism's temporal neutrality (but see the last section below). Earlier, we said that one could believe in prudence and its commitment to temporal neutrality without being committed to the egoist claim that all reasons for action are prudential. Viewed this way, Parfit's arguments that cast doubt on temporal neutrality are more general; they target prudence, as well as egoism.

Even if egoism is controversial, temporal neutrality is much less controversial. So in providing reductionist arguments against temporal neutrality Parfit is defending revisionary normative claims. He distinguishes more and less radical revisions (§102). The extreme claim concludes that the truth of psychological reductionism completely undermines the normative attitudes that are linked with personal identity. So, for instance, the extreme claim about special concern would imply that I have no special reason to be concerned about my own future. According to the moderate claim, the truth of reductionism does not completely undermine normative attitudes linked with personal identity, but may moderate those attitudes. For instance, in the case of future concern, the moderate claim does not reject all special future directed concern but does tolerate temporal discounting. For the most part, Parfit is agnostic between the extreme and moderate claims, but his sympathies seem to lie in the moderate response. He is also somewhat wishy-washy about whether the moderate should embrace these revisionary claims (e.g. temporal discounting) as rationally required or simply as not irrational, often insisting only on the weaker claim that such revisions are not irrational.

METAPHYSICAL DEPTH

Sometimes Parfit argues against prudence and its requirement of temporal neutrality by appeal to the claim that personal identity is not a deep metaphysical fact according to reductionism (309-12). Presumably, Parfit means that personal identity is not a mysterious further fact for the reductionist, as it is for the non-reductionist, but rather just consists in familiar physical and/or psychological facts and relations. But it’s not clear why this kind of metaphysical depth should have normative consequences. For instance, there is a comparable sense in which mental states are less metaphysically deep for a materialist than they are for a dualist. But that doesn’t make us think that the materialist should care less about people’s pain than a dualist.

Metaphysical depth alone doesn’t seem to have normative implications. Parfit needs to argue that specific features of psychological reductionism should force modifications in prudence’s temporal neutrality.
A DISCOUNT RATE

One such argument Parfit offers appeals to the way in which connectedness normally diminishes over time.

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 [313].

Notice a couple of things about this argument for discounting.

First, this is a discount rate with respect to psychological connectedness, not to time itself. Parfit's discount rate, therefore, should be distinguished from the discount rate with respect to time, which C.I. Lewis calls “fractional prudence.”3 Still, egoism and prudence must deny both kinds of discount rates.

Second, Parfit claims only that his discount rate is not irrational. Though temporal neutrality is inconsistent with this claim, it seems too modest. If the relations that concern should track hold to a reduced degree, then it seems that discounting should be rationally required.

How good is this reductionist rationale for intrapersonal discounting? One problem is that the argument assumes that special concern should track both connectedness and continuity (C&C), not just continuity (C). For it is only connectedness and not continuity that normally diminishes over time. But why suppose that the psychological reductionist's relation R should be formulated in terms of C&C, rather than C? Parfit and other psychological reductionists formulate their conception of personal identity in terms of continuity, not connectedness, as they must if they are to respond to Reid's transitivity worry.

It's true that Parfit argues that is relation R, rather than identity per se, that is the relation that matters (or has primary significance). But this conclusion was forced on us by duplication cases, which showed that an intrinsic relation, such as continuity, can deliver identity only by the addition of an extrinsic non-branching condition. But then the difference between relation R and identity should just be the addition of a non-branching clause. Those considerations give us no reason to revise our understanding of relation R in terms of psychological continuity.

It's worth thinking about some of the consequences of Parfit's reductionist rationale for discounting. As Parfit suggests, even though it becomes harder to view gross imprudence as irrational, it becomes easier to see it as a case of other-regarding harm involving immorality, instead of irrationality (§106). Is he suggesting that the shift from criticism as irrational to criticism as immoral involves no net change in our evaluation of gross imprudence? There could be a net change if, like Mill, we think it is easier to justify interference with conduct to prevent harm to others than it is to justify paternalistic interference. If Parfit is right, temporal discounting now falls within the scope of the harm principle.

Parfit's defense of the discount rate also seems to put severe limits on prudence. If special concern tracks connectedness, then there would seem to be a serious problem explaining why one has prudential reason to undertake improvements in one's character. For the greater the improvement, the more connectedness would seem to be reduced. But then the prudential rationality of character change is inversely proportional to the magnitude of the improvement. But surely self-improvement is one of the paradigms of prudential rationality.

Finally, it looks like the rationale for discounting doesn't go through, even if we assume special concern should track C&C. This is because Parfit's defense of the discount rate confuses parts with wholes. We can see how me-tomorrow is more connected with me-now than is me-in-

twenty-years. But the question is what I should do, and these different periods in my life are all equally parts of my life.

INTRAPERSONAL CONFLICTS OF IDEALS

Temporal neutrality can seem defensible when we restrict our attention to cases in which there is diachronic fixity of interests, because we can see how the agent is compensated later for the sacrifices she makes now. But what about cases in which there is significant change in an agent’s character or ideals?

In The Possibility of Altruism Nagel claims that temporal neutrality is unproblematic when "preference changes" are regarded with indifference. However, he sees a potential problem when neutrality is applied to intrapersonal conflicts of ideals.

It might happen that a person believes at one time that he will at some future time accept general evaluative principles -- principles about what things constitute reasons for action -- which he now finds pernicious. Moreover, he may believe that in the future he will find his present values pernicious. What does prudence require of him in that case? Prudence requires that he take measures which promote the realization of that for which there will be reason. Do his beliefs at the earlier time give him any grounds for judging what he will have reason to do at the later [time]? It is not clear to me that they do, and if not, then the requirement of prudence or timeless reasons may not be applicable [74].

Parfit shares Nagel's worries about the application of temporal neutrality to intrapersonal conflicts of ideals (155). Later, he describes the case of the nineteenth-century Russian nobleman.

In several years, a young Russian will inherit vast estates. Because he has socialist ideals, he intends, now, to give the land to the peasants. But he knows that in time his ideals may fade. To guard against this possibility, he does two things. He first signs a legal document, which will automatically give away the land, and which can be revoked only with his wife's consent. He then says to his wife, 'Promise me that, if I ever change my mind, and ask you to revoke this document, you will not consent.' He adds, 'I regard my ideals as essential to me. If I lose these ideals, I want you to think that I cease to exist. I want you to regard your husband then, not as me, the man who asks for this promise, but only as his corrupted later self. Promise me that you would not do what he asks' [327].

Parfit uses the Russian nobleman example to argue that adoption of a reductionist view of personal identity should lead us to revise our views about promissory fidelity, especially in cases involving intertemporal conflicts of ideals. But we can also use it to raise questions about the plausibility of the demands of temporal neutrality in such cases.

The Russian nobleman example is supposed to derive some of its force against the norm of temporal neutrality from Parfit's reductionist conception of personal identity. He seems to think that reductionism justifies the Russian nobleman's claim that that loss of his socialist ideals represents a substantial change, which he does not survive. This is what is supposed to justify the nobleman's wife in regarding his bourgeois successor as "another" who cannot revoke the nobleman's commitment.

But as our earlier discussion implies, there are several problems with this reductionist use of the Russian nobleman example. First, if this really were a substantial change, then prudence would not require neutrality between the socialist and bourgeois selves. Prudence requires intrapersonal neutrality but not interpersonal neutrality. If the example involves a substantial change, then it creates an interpersonal context. But then prudence does not demand concern and sacrifice for others. Absent some kind of reciprocation, these would be uncompensated sacrifices.
So if the change of ideals produced a substantial change, prudence would not counsel the nobleman to moderate his socialist ideals out of concern for his bourgeois successor. Second, psychological reductionism does not justify regarding the change of ideals in this case as a substantial change. Even if such changes of ideals disrupted psychological connectedness, they would presumably not disrupt psychological continuity. But reductionism needs to be formulated in terms of continuity, rather than connectedness, to avoid Reid’s transitivity concern. Moreover, psychological connections include ways an agent modifies his beliefs, desires, ideals, and intentions. So long as the nobleman plays a suitable role in generating and shaping his change of ideals (e.g. he is not the unwitting victim of psychological manipulation by another), character change of this sort is no obstacle to psychological connectedness. So in assessing the significance of the Russian nobleman example for the norm of temporal neutrality, we should resist any suggestion that socialist and bourgeois selves are literally different people. Both are equally parts of the nobleman’s life, and, as such, prudence demands temporal neutrality.

But avoiding Parfit’s reductionist gloss on intrapersonal conflicts of value does not itself remove the challenge that such cases pose to temporal neutrality. For we can still wonder if the demand of temporal neutrality makes sense in such cases. Should I be expected to moderate my pursuit of ideals I now hold dear for the sake of ideals I now reject but will or may later accept?

We should first notice something a little odd about the way intrapersonal conflicts of value are typically represented. Imagine that Before is at a crucial fork in the road of life and her prudential ideals speak in favor of route A, but she knows that she will later become After, whose prudential ideals will only be served if she now chooses route B. Should Before be true to her own ideals and choose route A, should she empathize with After and choose route B, or should she try to forge some third route C that compromises between A and B? This way of posing the problem assumes that there is a fact of the matter about the content of one’s future character and ideals independently of the crucial choices one makes now. But often, perhaps typically, this is false. One’s future character and ideals are very much influenced by crucial practical decisions one makes on the road of life. It is quite unlikely that a radical young socialist will turn into a complacent bourgeois regardless of the decisions he now makes. Who one becomes depends in part upon what one does now. But then it may be possible to avoid many intertemporal conflicts of value by making choices now that preserve, rather than compromise, one’s present ideals. Provided one’s present ideals are worthwhile (about which more below), one can honor temporal neutrality by acting in accord with one’s present ideals and thereby avoiding intertemporal conflict.

But perhaps some intertemporal conflicts are unavoidable. What then? Remember that prudence and its demand of temporal neutrality are, at least in the first instance, claims about what we have objective reason to do. The implications of temporal neutrality in situations involving unavoidable intrapersonal conflicts of value depend on the merits of the conflicting ideals. For present purposes, we can be quite ecumenical among different metaethical accounts of what makes some ideals meritorious and others meretricious. It will be helpful to divide unavoidable conflicts into symmetrical ones – those in which the merits of conflicting ideals are comparable – and asymmetrical ones – those in which the merits of conflicting ideals are very different.

The asymmetrical conflicts are perhaps more straightforward. There are two such cases. In the case of Corruption, Before’s ideals are valuable, whereas After’s are not. By contrast, in the case of Improvement, Before’s ideals are worthless, whereas After’s ideals are valuable. In cases of Corruption and Improvement, the demands of temporal neutrality are clear – act on the worthwhile ideal when you have it, not the worthless one. This is a claim about one’s objective reasons, the reasons one has in virtue of the facts about the situation whether one is in a position to recognize them or not. In these cases, temporal neutrality does not require neutrality between current and future ideals.

It is fairly easy to see how the agent can and will act on these reasons in the case of Corruption, for this just requires acting on his current ideals. Here acting on one’s current ideals is
also what temporal neutrality requires.

However, matters are more complicated in the case of Improvement. Temporal neutrality’s claim about one’s objective reasons remains plausible. One has objective reason to act later on those valuable ideals that one will hold, rather than the worthless ideals that one now embraces. But can one act on this verdict if it is the worthless, rather than the valuable, ideal that one now embraces? Can temporal neutrality make plausible claims about subjective rationality? Could it be subjectively rational to act on valuable ideals that one does not now hold? The answer is Yes, provided that we understand subjective reasons as the reasons one has, not in virtue of what one now judges, but in virtue of what it would be reasonable for one to judge now if one gave the matter due attention. It is part of a theory of subjective rationality to specify more precisely what kind of idealization of the agent’s epistemic situation is appropriate in determining her subjective reasons. As long as the worthlessness of Before’s ideals and the merit of After’s ideals do not transcend reasonable idealizations of the agent’s epistemic situation, whatever those are, the comparative merits of earlier and later ideals will be ascertainable in the relevant way. If the comparative value of her current and future ideals is available to her in this way, we can ascribe to her a subjective reason to favor her future ideals. However, in cases of Improvement in which the comparative values of current and future ideals is a transcendent fact (transcending the relevant idealization), then the demands of objective and subjective rationality appear to diverge. The friend of prudence can and should defend temporal neutrality as a claim about the agent’s objective reasons. Whether she is in a position to recognize it or not, she has no reason to act on her current ideals and will have reason to act on her future ideals. This can be a case where it may not be subjectively rational to do what is in fact objectively rational.

What about unavoidable conflicts whose merits are symmetrical? The Minus-Minus situation occurs when the conflicting ideals are similarly worthless. Here it seems right to agree with neutrality’s claim that there is objective reason not to act on either ideal but to find, adopt, and act on some third ideal that has merit. Provided that the comparative merits of the meretricious and genuinely valuable ideals are reasonably ascertainable and are not (in the relevant sense) transcendent facts, this also yields a plausible claim about the agent’s subjective reasons. The agent should act on neither meretricious ideal but adopt and act on the new valuable ideal.

Perhaps the most interesting case of unavoidable conflict is the symmetrical case in which the conflicting ideals are both valuable and comparably so. One example might be a conflict between excelling as a professional athlete early in life, which may require forgoing extended educational and professional training and may impose significant health costs later in life, and various forms of professional and personal success later in life. Another example might be familiar conflicts between success in professional and family life. We might call any such case a Plus-Plus case. By hypothesis, the conflict is unavoidable, so that After’s ideals conflict with Before’s no matter what the agent now does, and each ideal is valuable. Here, temporal neutrality recognizes a conflict of objective reasons and counsels a kind of neutrality among the competing ideals. On reflection, this seems right. If the agent can pursue Before’s ideals unreservedly only by completely frustrating After’s ideals (and vice versa), then there seems something objectively wrong with the unreserved pursuit of present ideals. Ideally, one would try to find a way to achieve substantial success in one’s ideals both now and later, even if it required some moderation in or restrictions on the pursuit of one’s ideals now or later. Neutrality’s counsel of moderate or restricted pursuit of current ideals is an instance of the familiar adage “Not to burn one’s bridges”. Where such compromise and accommodation are possible, neutrality makes good normative sense. Call these cases of Accommodation. But accommodation may not always be possible. In cases of Genuine Dilemma there is no prospect of substantially accommodating both ideals. Here, neutrality seems compatible with two possibilities. On the one hand, one might achieve some less-than-substantial success along both ideals – neither a stellar success nor an abject failure at any time. Alternatively, one might engage in the unreserved and successful pursuit of ideals either now or later (but, by
hypothesis, not both), provided that the process of selecting the favored ideal gave equal chances of success to both ideals (as in a coin flip). Neither alternative is attractive, but that seems to be a consequence of the situation being dilemmatic. One consolation is that unavoidable conflicts are somewhat rare, and Genuine Dilemmas are even more exotic. Neutrality's claims about our objective reasons in such cases seem plausible enough. And, as before, provided the merits of the conflicting ideals are not transcendent facts, these claims about the agent's objective reasons apply to her subjective reasons as well.

We can reconcile the demands of prudence and fidelity to one's ideals if we remember that the agent is a person who is temporally extended. Her past, present, and future are equally parts of her and her life. To be true to herself, since she is a temporally extended person, she must be true to all of her reasonable ideals and cannot be selectively attentive to her current ideals. She must weigh her future reasonable ideals, where these are fixed, against her current reasonable ideals, where this is necessary, in order to conform her behavior to all of her reasonable commitments. This sort of concern for one's whole life does not require forsaking one's current prudential ideals. But it does require conditioning their pursuit on recognition of the legitimate claims that one's reasonable future prudential ideals make on one.

REDUCTIONISM AND AGENT BIAS

Though Parfit's critical focus in ch. 14 is on egoism's temporal neutrality, he also has the resources to challenge egoism's agent bias. Fission implies that concern should track psychological continuity, rather than identity. But fission involves interpersonal psychological continuity. Tom should be just (or nearly) as concerned for Dick and for Harry as he would be for himself in non-branching circumstances. But this intrinsic concern for anyone to whom one is R-related seems to conflict with egoism's agent bias, for in fission Tom appears to have non-derivative reason to be concerned about someone other than himself. Can we reconcile egoism with fission?

We could accept the claim that interpersonal continuity justifies concern and yet defend egoism by reformulating agent bias in terms of continuants (maximal R-related selves), rather than persons. In fission there are three persons, but only two continuants, who share a person and body in Tom. On this view, we treat continuants, rather than persons, as the relevant agents.

Alternatively, we could accept the claim that interpersonal continuity justifies concern and yet defend egoism by claiming that psychological continuity extends the agent's (the person's) interests, much as Plato claims that interpersonal love is the next best thing to immortality. As Plato and others recognize, it is common in intimate associations to say that the beloved is an extension of the lover and that the beloved's interests are a part of the lover's interests. It's plausible that the sort of psychological interaction and influence we find in intimate associations is the basis for these claims. But interpersonal psychological continuity is maximal in the fission case and so is the strongest case for saying that one person is an extension of another and the interests of the second are part of the interests of the first. On this view, psychological continuity extends one's interests, and because interpersonal association is a form of interpersonal psychological continuity, interpersonal associations extend one's interests in proportion to the strength of the association. This yields an extended conception of self-interest to which the egoist can appeal. Indeed, there's some reason to think that the central claims of self-referential altruism could be represented as theorems of egoism. But these are issues to which we will return later.