Comments

Derek Parfit

CONSEQUENTIALIST RATIONALITY

What is personal identity? What is involved in the continued existence of the same person over time? After defending one view about this subject I argued that, on this view, it is irrational to care about personal identity in the way that most of us do (pp. 199–280).[1] In her elegant and thoughtful paper, Susan Wolf claims that what it is rational to care about has "nothing to do with the metaphysics of personal identity." Whether we "rationally ought" to care about identity depends on whether, if we do, this will be better for us (Susan Wolf, "Self-Interest and Interest in Selves," in this issue, p. 713).

Wolf's Consequentialism seems to me to go too far. As I wrote, if some desire has good effects, this fact cannot show that this desire is rational; it can at most show that we have a reason to try to have, or to keep, this desire. It would be irrational, for example, not to care about future Tuesdays. If something will happen on a Tuesday, this is no reason for caring about it less. But, if we shall have to endure weekly ordeals, and could schedule these for Tuesdays, it might be better for us if we had this pattern of concern. This would give us a reason to try to become in this respect irrational (pp. 7–13, 169).

Consider next our bias toward the future: the fact that, when we think about our lives, looking forward affects us more than looking backward. If we lacked this bias we would be more like Timeless, my imagined temporally neutral man. This man's attitude to time is, in some ways, bad for him, since he is not glad when bad things are over. But he is also not sad when good things are over. One consequence is this. As Timeless grows older, though he has less and less to look forward to, he has more and more to look backward to. Even when he is about to die he remains serene since, though he then has nothing to look forward to, he can look backward to his whole life.

If we were like Timeless, we would be less depressed by aging and the approach of death. This and other good effects would, I argued, outweigh the bad effects. If we were temporally neutral, this would be,

1. References in parentheses are either (as here) to my book *Reasons and Persons* (Oxford: Clarendon Press, 1984) or to one of the six preceding articles in this issue. In writing these comments I have been greatly helped by Janet Radcliffe Richards.

Ethics 96 (July 1986): 832–872

on the whole, better for us. This fact gives us a reason to try to become temporally neutral; but it does not show that this attitude is rational. Though Timeless is, on the whole, better off than we, it may be *he* who is irrational. Thus, he is not relieved when some ordeal is over. "Why should I be?" he asks. "The ordeal is just as painful, and just as much a part of my life. What difference does it make that it is now over?" Many people think that, when his ordeal is over, Timeless *ought* to be relieved. Temporal neutrality seems to them crazy—to involve an absurd mistake. Whether this is so depends, I claimed, on the nature of time. Thus if time's passage is a myth, as some philosophers and scientists believe, temporal neutrality cannot be irrational (pp. 165–86).

I made a similar claim about our attitude to personal identity. On my view, the rationality of this attitude depends, not on its effects, but on the nature of personal identity. Wolf claims the opposite. On her view, the rationality of this attitude depends entirely on its effects. I agree that, if some attitude has good effects, we have a reason to try to have this attitude. But Consequentialism is not the whole truth about rationality. Whatever the effects, it would be irrational not to care about future Tuesdays.[2]

EGOISM AND THE FEAR OF DEATH

In ways that I discuss below, most people care a great deal about personal identity. I argued that on the true view about identity, which I called "Reductionist," this attitude is irrational. Wolf predicts that, if we ceased to have this attitude, this would be worse for us. If Wolf's prediction was correct, this would not show that this attitude is rational; but it would give us a reason to try to keep this attitude. And, if this attitude would be undermined by our coming to believe the Reductionist View, we would have a reason to deceive ourselves.

Is Wolf's prediction correct? On the Reductionist View, what it is rational to care about are various psychological connections. Wolf claims that, if these connections were what we cared about, this would affect our concern about our own future. We would care less about those parts of our future to which we are less closely psychologically connected (Wolf, p. 710). I agree that this may happen, and that it may have bad effects. It may lead the young, for example, to impose burdens on the old people whom they will become.[3]

Most of Wolf's claims are about our relations with other people. She suggests that, if we ceased to care about identity, our loves and friendships

2. Wolf's n. 4 qualifies her Consequentialism, as does her account of why persons are especially valuable. But she also believes that her arguments "are not simply at odds with Parfit's . . . they undermine his arguments . . . completely" (Wolf, p. 714). This could not be so unless Consequentialism was the whole truth about rationality.

3. Acting in this way, because we have this pattern of concern, cannot—I argued—be claimed to be irrational (except in extreme cases). But it may be immoral. We ought not to do to our future selves what it would be wrong to do to other people (pp. 307–20).

would be "shorter lived," since they would be more easily destroyed by changes in the people whom we love (Wolf, p. 713). This too may happen. But I do not believe that, as Wolf suggests, our loves and friendships would be "shallower." Even short rivers may be deep. Nor do I believe that these relations would be based "exclusively on . . . merit" (Wolf, p. 719). Wolf's claims are about the effects of *changes* in the people whom we love. Though such changes may remove the causes of this love, they do not affect what these causes are.[4]

I have questioned some of Wolf's claims; but I accept others. As she predicts, if we became Reductionists, this might have some bad effects.[5] But there would be other effects, some of which would be good. And the good effects would, I believe, outweigh the bad. I believe that, if we became Reductionists, this would be on the whole better for us.

Wolf's predictions appeal to the claim that psychological connections may be, over longer periods, weaker. Since this claim is obviously true, it is not the important part of the Reductionist View. What is important is not what this view claims but what it denies. As I argued, even if we are not aware of this, most of us have certain strong beliefs about the nature of personal identity. On the Reductionist View, these beliefs are not true. It would be the loss of these beliefs which would have the main effects on what we care about and what we do.

Though these beliefs apply to actual cases, they are best explained with the help of imaginary cases. I describe such cases here not to challenge our beliefs but merely to suggest what these beliefs are—to suggest how, if we became Reductionists, our beliefs would change.

One imaginary case is what I called "Teletransportation." If this happened to me, a Scanner would destroy my brain and body, while recording the exact states of all my cells. A Replicator would then create, in another place, a brain and body just like mine. To some readers of science fiction, Teletransportation seems to be a way of traveling; others think it a way of dying.

Consider the variant which I called the "Branch-Line Case" (pp. 199–201). Because the New Scanner does not destroy my brain and body, I am able to meet my Replica. While we are talking, I learn that I am about to die. My Replica tells me not to be concerned. Since he is exactly like me, he will take up my life where I leave off. He will look after my

4. Wolf's discussion of love seems to me both moving and convincing. As she argues, "We do not want or expect reason to govern here . . . we can give reasons why it is good *that* particular persons matter to each and perhaps also why it is good that they do not look for reasons for mattering" (Wolf, pp. 719, 719–20). It is because most love is not based on reasons that it would be little affected by a change in our view about the nature of personal identity.

5. I assume here that, if we became Reductionists, we would cease to care about identity. Wolf may seem to be an exception: a Reductionist who still cares about personal identity. But, as I suggest in Appendix B below, identity is not in fact what Wolf cares about.

children and complete my book. And no one else will ever know that he is not me.

Should I be concerned? As a Reductionist, I would say: "My prospect is as good as ordinary survival. What it is rational to care about are the various psychological connections which, in ordinary cases, unify a person's life. Since I am connected to my Replica in all these ways, I have no reason for concern." But, though I accept the Reductionist View, I would find these claims hard to believe. It would be far easier to believe that what matters is that, soon, there will be no one living who is me. On this more natural view, though it may be some consolation that I shall have a Replica, my prospect is nearly as bad as ordinary death.

In the Branch-Line Case it is fairly clear what will happen to me. I described several cases where this would be less clear (pp. 229–306). One example is the range of cases which I called the "Physical Spectrum." In each of these many cases, a future person would have some proportion of the cells in my brain and body. This proportion would, in the different cases, range from *all* to *none*. Since my other cells would be replaced with duplicates, the resulting person would in each case be just like me.

In a case in the middle of this range, what should I expect to happen? Suppose I know that before tomorrow half my cells—or three-quarters or nine-tenths—will be replaced with exact duplicates. The natural question is "Would the resulting person be *me*, or would he merely be *someone else* who is just like me?"[6]

A Reductionist would claim: "This is an *empty* question. These are not two different possibilities, either of which might be true. They are merely two descriptions of the same course of events. When you know which of your cells will be replaced, you know everything." Once again, I would find these claims hard to believe. How can I know everything *when I do not even know whether I shall live or die?* If we imagine being in my place, most of us would react in the same way. We might say, "Any future person must be either, and quite simply, me or someone else. If there will be someone tomorrow who is in pain, either I shall feel that pain, or I shan't. One of these must be true."[7]

6. Some people believe that, in all these cases, the resulting person would be me. If we believe this, we should consider what I called the "Combined Spectrum." As in the Physical Spectrum, in each of the cases in this range a different proportion of my cells would be replaced. But the new cells would not be duplicates. If the resulting person would have fewer of my cells, he would therefore be less like me. At the far end of this spectrum the resulting person would be just like Greta Garbo. There would be no physical connection, and few similarities, between her and me. It is clear that, in this case, the resulting person would not be me. My remarks below about the Physical Spectrum can be reapplied, with certain changes, to the cases in the middle of this Combined Spectrum (see my pp. 234–43).

7. That this is the natural view is argued in B. Williams, "The Self and the Future," in *Problems of the Self* (Cambridge: Cambridge University Press, 1973); see Appendix A below.

On the Reductionist View, our identity over time just consists in various physical and psychological connections. In the imaginary cases I have just described, the physical connections would hold to different degrees. In other cases, some of them actual, it is the psychological connections which would be matters of degree (pp. 227–33, 236–39). As our reactions to such cases show, most of us are not Reductionists. We do not believe that our continued existence merely involves such connections. It seems to us to be a further fact, of a deep and simple kind: a fact which, in every case, must be either wholly present or wholly absent. *There is no such fact.* This is the important part of the Reductionist View.

If we accept this view, as I argued that we should, this may affect what we care about and what we do. As Wolf suggests, there may be some changes in our relations with other people. But these would not be the main effects. It is when we think about ourselves that the Reductionist View is hard to believe, and this is where the main effects would come.[8]

We have two kinds of concern about our own future.[9] One is *direct* concern, such as our fear or dread of pain and death. The other is *derivative* concern: a concern about ourselves which results from having various other concerns. We want, for example, to remain active so that we can achieve certain ambitions, and protect those whom we love. If we became Reductionists, this would not affect our derivative concern; but it would affect our direct concern.

Reconsider the Branch-Line Case. I claimed that, since the Reductionist View is true, my relation to my Replica is as good as ordinary survival (pp. 282–89). This claim can be reversed. Ordinary survival is as bad as, or no better than, being destroyed and replicated. What we fear will be missing, after we die, is always missing. Our survival never involves the specially intimate relation in which we are inclined to believe.

If we grasped these truths, we would care less about our own future. We would have for ourselves in the future only the concern that we would have for a mere Replica. Suppose I know that tomorrow I shall be in pain. If I knew that, after my death, a Replica of me would be in pain, I would not fearfully anticipate this pain. And my relation to myself tomorrow is no closer than my relation to my Replica. It is hard to grasp this truth. When I forget the arguments, my belief in the further fact returns. But when I reconvince myself, this for a while stuns my direct concern.[10]

On the Reductionist View, since there is no further fact, it is irrational to regard personal identity—or our own continued existence—as what matters.[11] What it is rational to care about are the various psychological

8. Another kind of effect, on our moral beliefs, I discuss below.

9. See J. Perry, "The Importance of Being Identical," in *The Identities of Persons,* ed. A. Rorty (Berkeley: University of California Press, 1976).

10. Can it be permanently stunned? Some Buddhists may have found the answer.

11. See also Appendix B below.

connections which, in this imagined case, would hold between me and my Replica.[12] Wolf predicts that, if we ceased to care about identity, our affection for other people would become more fragile. Even if there were such bad effects, I believe that they would be outweighed by the lessening of our concern about ourselves. Our natural egoism is often bad for others, and it makes our own lives bleaker.[13]

Wolf makes another prediction. If we ceased to care about identity, we might "aspire to and accomplish less." We might try to avoid any major psychological change, because such a change would seem in advance like "an early death" (Wolf, p. 712). But such changes do not seem to me like death. Indeed, when it is better described, even *death* does not seem like death. Instead of thinking, "I shall die," I should think, "After a certain time, none of the experiences that occur will be connected, in certain ways, to these present experiences." In this redescription my death seems to disappear.

PERSONAL IDENTITY AND INJUSTICE

I argued that, if the Reductionist View is true, this supports the rejection or revision of various claims about just punishment and fair distribution. In his lucid paper, Bart Schultz questions these arguments (Bart Schultz, "Persons, Selves, and Utilitarianism," in this issue).

One of my arguments was this. On the Reductionist View, the fact of personal identity is less deep or involves less. Because this fact is less deep, it is more plausible to deny that this fact is morally important. Since distributive principles assume this fact to be morally important, it is more plausible to reject these principles. More exactly, the rejection of these principles is more plausible than it would be on the Non-Reductionist View (pp. 329–42).[14]

12. According to some Reductionists, it matters whether these connections have their normal cause: the continued existence of our brains. I do not see how this could matter, unless there is a further fact which an abnormal cause would fail to produce. But my discussion of this question (pp. 282–97, 468–77) seems to me now inadequate. (I should not have written "any cause" in my claim about identity on p. 216. This suggests that, if my brain and body were destroyed before the creation of my Replica, this Replica would be me. My main claim is that, in such cases, the question about identity is empty.)

13. I have claimed that, though it would be bad if we cared less about our further future, it would be good if we cared less about ourselves. These claims are not, as they may seem, inconsistent. If we care less about our further future, we have a less impartial pattern of concern. We are thus more likely to do what, impartially considered, has bad effects. If we care less about ourselves, we have a more impartial pattern of concern, since we are less biased in our own favor. This would have good effects. (It would be bad, however, if we were all wholly impartial [pp. 27–30].)

14. In her comments on this argument, Wolf claims that nothing is shown by my appeal to the analogy between persons and nations (Wolf, pp. 717–18). I agree. She also challenges my claim that, if we move from the Non-Reductionist to the Reductionist View, "It becomes more plausible to be more concerned about the quality of experiences, and less concerned about whose experiences they are" (p. 346). Wolf writes, "The value of

This argument does not show that we ought to reject these principles. Schultz therefore claims that it may give *no* support to their rejection (Schultz, p. 740). But a gain in plausibility is not nothing. And such a gain may be great even though, as Schultz says, it cannot be *proved* to be more than "marginal." We should not assume that, to support a moral belief, an argument must be decisive.

Schultz's main objection is to the vagueness of my claim that, on the Reductionist View, personal identity is "less deep." I accept this objection. As Schultz writes, if we argue that Reductionism supports or undermines some moral principle, we should try to explain "how and why" this is so. We should try to describe "the specific connection between the fact of personal identity and the principle in question" (Schultz, p. 732).

Two of my arguments did just this (pp. 324–25, 342–45). I shall here revise and extend these arguments.[15] Suppose that, in the Branch-Line Case, I had earlier committed some crime. When I talk to Backup—as my Replica is called—I warn him to escape. But he is caught and convicted. The judge says, "Given the gravity of Parfit's crime, you deserve a life sentence. Though you are not Parfit, between you and him there are all of the normal psychological connections. You have apparent memories of Parfit's life, and in every other way you resemble him. These connections are enough to make you guilty." Backup protests, "This is outrageous. These connections are irrelevant. I did not choose to resemble Parfit, or to have these apparent memories. I cannot deserve to be punished for what Parfit did before I even existed."[16]

Most of us would side with Backup. We would believe that, in the absence of personal identity, these psychological connections cannot carry with them desert or guilt. But on the Reductionist View personal identity merely consists in these connections. Backup is not me only because, in this case, these connections do not have their normal cause: the continued

---

persons is not, as this proposal would suggest, dependent on their ability to have such momentary experiences" (Wolf, p. 709). But I was not discussing the value of persons. I was discussing the relative importance of the *amount* of suffering that is suffered and the *distribution* of this suffering between different people. My claim was that, as Reductionists, we may care more about the *nature* of what happens, and care less about *who* the persons are to whom it happens (p. 340). Wolf also writes that the Reductionist View should at most affect our answer, not to the question of why it matters to whom something happens, but to the question why it matters that what happens, happens to a person at all. Theories about personal identity are at most relevant, she claims, not to the distinction between different persons, but to the distinction between persons and other entities (Wolf, p. 718). This claim puzzles me. When I discussed the distinction between persons, I was asking why it matters whether benefits and burdens come to the same or different people. This is a question about the importance of personal identity. I do not see why theories about personal identity cannot be relevant to this question.

15. These arguments are partly due to M. Wachsberg's excellent "Personal Identity, the Nature of Persons, and Ethical Theory" (Ph.D. diss., Princeton University, 1983).

16. Backup would not protest if he mistakenly believed that he was me. But the important question is what, if he knew the truth, he could justifiably claim.

existence of my brain.[17] Is it the absence of this normal cause which makes Backup innocent? Most of us would answer no. We would think him innocent because he is not me.

This reply would show that we are not Reductionists. The fact that Backup is not me seems to us to be different from, and more important than, the fact that the psychological connections have an abnormal cause. What we believe to be missing is not the normal cause but the further fact: the specially intimate relation which we assume to be involved in our own continued existence over time. This is the fact which, on our view, carries with it desert and guilt.

Suppose next that we become Reductionists: we decide that there is no such fact. An obvious conclusion follows. If it was only this fact which could carry with it desert and guilt, these have also disappeared. No one ever deserves to be punished for anything they did.[18]

Beside this argument about desert, I gave a similar argument about principles of just distribution. According to these principles, many resources ought to be fairly shared between different people. Like the principle of desert, distributive principles assume that personal identity is morally important. When our acts will affect only ourselves, we can ignore these principles. In accepting a burden for the sake of a later benefit, we cannot be treating ourselves unfairly. Why is this not possible? Because the burdens that we bear can be fully *compensated* by benefits at other times. Such compensation requires personal identity: the burdens that *we* bear cannot be compensated by benefits to *other people*. It is therefore claimed to be unfair when we bear the burdens and others receive the benefits. On this view, it is the nonidentity of different persons—or what is sometimes called "the separateness of persons"—which supports the claim for fair shares.[19] (These remarks extend our ordinary use of the word "compensation." This applies to cases where, because we have been harmed, we are later benefited. On the extended use, we can be compensated *in advance*, and for any kind of burden.)[20]

17. I assume that it does not make a difference that my life and Backup's briefly overlap (see my pp. 266–71, 287–89).

18. The Reductionist View does not, by itself, imply this conclusion. But this is the conclusion we should draw if we continue to believe that Backup cannot deserve to be punished for my crime. (It is worth adding a detail to the case. Suppose that, because I knew that Backup would not be punished, I deliberately arranged to be destroyed and Replicated. Since I thought my relation to Backup to be no worse than ordinary survival, this seemed to me a costless way to evade punishment. When discussing a similar case, Wiggins wrote, "A malefactor could scarcely evade responsibility by contriving his own fission" [quoted (and inconsistently endorsed) on my p. 271]. But we may still believe that Backup cannot be guilty, since *he* did not *choose* to be my Replica.)

19. See J. Rawls, *A Theory of Justice* (Cambridge, Mass.: Harvard University Press, 1971), pp. 26–29; R. Nozick, *Anarchy, State, and Utopia* (Oxford: Basil Blackwell, 1974), p. 33; and T. Nagel, *The Possibility of Altruism* (Oxford: Clarendon Press, 1970), pp. 138–42, and *Mortal Questions* (Cambridge: Cambridge University Press, 1979), pp. 124–25.

20. We can receive and spend, before we earn it, what some employers call our "compensation."

Return now to the Branch-Line Case. Suppose that, in the past, I took more than my fair share of society's resources. Backup is therefore given less than a fair share. He is told, "Though you are not Parfit, between you and him there are all of the normal psychological connections. You have apparent memories of Parfit's life, and you are just like him. These connections make it fair that you should have few resources. For the poverty which you must endure Parfit's luxury gave you, in advance, full compensation."

Backup again protests, "This is unjust. The psychological connections are irrelevant. My apparent memories of Parfit's luxury merely make my poverty harder to bear. *I* was not compensated by what *he* enjoyed before I even existed."

Once again, most of us would side with Backup. What would our reason be? What makes it impossible for Backup to be compensated by my luxury? Is it the fact that the psychological connections have an abnormal cause? Most of us would answer no. We would believe that Backup was not compensated because he is not me. This reply would again show that we are not Reductionists. The fact that Backup is not me seems to us to be different from, and more important than, the fact that the psychological connections have an abnormal cause. What we believe to be missing is not the normal cause but the further fact. On our view, this must be the fact which, within a single life, makes compensation possible.

Suppose next that we become Reductionists, deciding that there is no such fact. We should therefore draw another conclusion. Since it was only this fact which made compensation possible, it is never possible.

What does this imply? Nagel writes, "The criteria of personal identity . . . determine the size of the units over which a distributive principle operates."[21] These units are, he assumes, people's lives, since within a life—or when they all come to the same person—burdens can be compensated by greater benefits. On the argument just given, the relevant units cease to be people's lives. Since there cannot be compensation over time, the relevant units are the different states which, at different times, people are in. We should therefore aim for a fair distribution, not between different people, but between the different moments within each life.

This argument raises many questions. Does it, for example, support the Utilitarian view? This is unclear. According to this argument, we should greatly extend the scope of distributive principles. Since Utilitarians reject these principles, the argument may take us even further from the Utilitarian view. But, as I remarked, a change in *scope* may justify a change in *weight*. When we extend distributive principles in this extreme way, they may become less plausible. We may decide to give them either less

21. Nagel, *Mortal Questions*, pp. 124–25, n. 16.

or even no weight. The argument may thus *indirectly* support the Utilitarian view.

If we still give weight to distributive principles, we are rejecting this view. But the position we have reached may be, in its implications, similar. Consider the principle which claims that we ought to give priority to helping the people who are worst off. If the relevant units for distributive principles are people's states at particular times, what corresponds to those who are worst off are the worst states that, at particular times, people are in. On this principle, we should give priority to making these states less bad. This is what Negative Utilitarians—those who give priority to relieving suffering—claim that we should do.

Schultz says little about the argument which I have just described; but in discussing another question he makes some relevant remarks. He asks whether Reductionism could *directly* support the Utilitarian view. For this to be so, he writes, Reductionists would have to answer "the Rawlsian objection to Utilitarianism." They would have to show that persons do not have "those characteristics" which make this objection "seem compelling" (Schultz, p. 732). And Schultz claims that this could not be shown.

This claim seems correct. Utilitarians aim for the greatest net sum of benefits minus burdens, whatever their distribution between different people. The Rawlsian objection is that burdens to one person cannot be compensated by benefits to someone else. This objection appeals to the fact that these benefits and burdens come to different persons. This is the fact which makes it impossible for there to be compensation over different lives. Since Reductionists do not deny this fact, they cannot directly answer the Rawlsian objection.

The argument described above does not claim to answer this objection. It extends the scope of this objection, claiming that, even within the same life, there cannot be compensation over time. For this to be a good argument, it must appeal to the fact which makes the Rawlsian objection "seem compelling." It must show that, on the Reductionist View, this objection ought to be applied even to the different parts of the same life.

The Rawlsian objection appeals to the separateness, or nonidentity, of different persons. Which is the relevant feature of this nonidentity? Is it the absence of direct psychological connections between the lives of different persons? If this is why there cannot be compensation over different lives, Reductionists could not conclude that, even within the same life, there cannot be compensation over time. They admit that, between the different parts of the same life, there are direct psychological connections.

There is an obvious way to approach this question. Nonidentity normally involves the absence of direct psychological connections. But we can imagine cases where these connections hold between two different

people. In such a case, would benefits to either of these people compensate the other's burdens? Would this be true, for example, of me and Backup?[22]

Schultz writes, "My replicas can, it would seem, still receive compensation for harm done to me" (Schultz, p. 737). But this is not the answer. The question is not, Can Backup be compensated for my burdens? The question is, Can benefits to Backup compensate *me*? And, if Backup bears a burden, was *he* compensated in advance by benefits to me?[23]

Most of us, I claimed, would answer no. We would believe that, despite the psychological connections between me and Backup, he was not compensated by my luxury. And we would believe this, not because of the abnormality of the cause of these connections, but because Backup is not me. This would show that we believe in the further fact, and believe that only this fact makes compensation possible. The rest of the argument is brief. Since there is no such fact, compensation over time is not possible. We should apply the Rawlsian objection even to the different parts of the same life. The units for distributive principles become the states that people are in at different times.[24]

Schultz's main claim is that "it is still unclear" whether Reductionism "could support . . . [a] moral theory in the way that Parfit suggests" (Schultz, p. 722). He asks me to explain "how and why" this can be so. The arguments described above provide this explanation. What they appeal to is precisely what Reductionists deny. These arguments claim that only the further fact could carry with it desert and make compensation possible. When we consider my imagined case, most of us would accept these claims. We would believe that Backup deserves no punishment, and that he was not compensated by my luxury. If we believe this, we should conclude that no one deserves to be punished, and that there is never compensation over time.

These two arguments are stronger than the argument I gave for the Utilitarian view.[25] That argument was vague, and it could not show that

22. As M. Herrman has suggested, another way to approach this question is to consider cases in which, in the life of the same person, various psychological connections do not hold. (Some of what I wrote implies that, if there were none of these connections, we could not still have the same person. It would be better to claim that this is a "degenerate case": a case in which though there is identity, there is not what matters.)

23. Schultz may have in mind what Appendix H below calls "quasi-compensation." But the point just made still stands.

24. Appendix H qualifies this conclusion.

25. On my pp. 329–42. (Schultz discusses a reply by Rawls to an earlier version of that argument. Rawls claimed that the degree of psychological connectedness within each life varies in different societies, and that in a society of Kantians there would be enough connectedness to support the requirements of distributive justice. Schultz points out that, since my argument appealed not to a claim about connectedness but to a claim about depth, this reply needs to be restated. He suggests that this can be done; in a society of Kantians, he writes, the unity of people's lives may be deeper (Schultz, p. 743). But this reply cannot be restated so that it answers the different argument described above. A society of Kantians could not produce the further fact.)

we ought to accept this view. These two arguments are less vague, and they may show that we ought to accept their conclusions.[26]

AGAINST THE SELF-INTEREST THEORY

On the Self-interest Theory, or S, what each of us has most reason to do is whatever would be best for himself, and it is irrational for anyone to do what he believes will be worse for himself. Shelly Kagan challenges my arguments against this theory (Shelly Kagan, "The Present-Aim Theory of Rationality," in this issue).

The central claim of S is

> S1: For each person, there is one supremely rational aim: his own self-interest.

According to S, it is irrational not to be governed by this aim. Even if we care more about helping those we love, or saving people's lives, or preventing injustice, it would be irrational to achieve these aims at the slightest cost to our own self-interest.

My first argument was this.[27] We should reject S1. There are many aims which are no less rational than the aim singled out by S. Since these other aims are no less rational, it would be no less rational to act upon them. This would be no less rational even when the agent knows that, in achieving one of these other aims, he is acting against his own self-interest. These claims form part of what I called the Present-aim Theory, or P (pp. 117–33).

Kagan objects that this argument "begs the question against S" (Kagan, p. 752). The central question is not, he claims, about the rationality of different aims: it is about the relation between rationality and time. He writes, "The issue between S and P is the acceptance or rejection of temporal neutrality. . . . For S, time is irrelevant; for P, it is paramount. . . . An adequate defense of the Present-aim Theory will need to [explain] why temporal relativity is appropriate in a theory of rationality. In particular, it will need to explain what it is about currently held desires that allows such desires—and only such desires[28]—to generate reasons directly. It is the absence of such an account, I believe, that dooms Parfit's defense of the Present-aim Theory" (Kagan, pp. 749, 750, 759).

26. I should not have written that we could defensibly deny these conclusions (pp. 325, 343). I do not know whether this is possible. (Schultz writes that, if Reductionism shows that there cannot be compensation over time, "Some type of utilitarianism would be on the whole . . . appropriate. . . . But again, given the general vagueness about just what follows from the Reductionist View, it is impossible to say whether utilitarianism actually could receive *any* theoretical support from it" [Schultz, p. 741; my italics]. This again seems to assume that, to support a moral claim, an argument must be decisive.)

27. This "argument" hardly deserves the name, since it is little more than a question. My main aim was to make this question clear and explain why it is important (pp. 117–36).

28. See n. 35 below.

Kagan shows that I was confused. I did suggest that the central question is whether rationality requires us to be temporally neutral.[29] But this is *not* the question raised by my first argument. That argument claims that self-interest is not the one supremely rational aim: many other aims are no less rational. This claim provides what I called the best objection to the Self-interest Theory. And this claim does not even *mention* time.

Since this claim does not mention time, it does not show that we ought to accept the Present-aim Theory. As Kagan saw, there is a gap in my defense of P. But this does not "doom" this defense, since this gap is easily filled.

Kagan writes:

> Parfit's first argument comes to this: S is necessarily intolerant in its attitude toward the rationality of different patterns of concern. It implausibly elevates one particular pattern—the bias in one's own favor—and gives it a unique theoretical status. The argument against S is thus fairly direct. On the other hand, the support this argument provides for the Present-aim Theory is relatively indirect. It is clear, however, that Parfit believes that P can be tolerant where S must be intolerant: P can accept the rationality of numerous patterns of concern, and instruct the agent to act in conformity with whatever pattern happens to reflect his strongest desires. [Kagan, p. 750]

This argument for P is indirect because, in accepting the rationality of numerous patterns of concern, we are not accepting P. To reach P we must add the claim that, if someone has one of these equally rational patterns of concern, *this* is the pattern of concern on which it is rational for *this person* to act.

Though P needs this extra claim, it is hard to deny. If there are many equally rational aims, what it is rational for *me* to do must depend on which, among these many aims, are *my* aims. And what it is rational for me to do *now* must depend on which, among these aims, are my aims now. As Kagan writes, if we accept a *pluralistic* view about the rationality of different aims, "we have reason to reject S and to accept P" (Kagan, p. 750).

Suppose, for example, that we accept Sidgwick's minimally pluralistic view.[30] Suppose we believe that there are two supremely rational aims: what would be best for ourselves and what would be morally best. On this view, when morality conflicts with self-interest, neither provides the stronger reason for acting: it would be rational to follow either. But would it be rational to follow either, even if we only cared about the other? Would it be rational for a criminal to surrender to the police, even

29. Thus I called this part of my book "Rationality and Time," and my section 52 misdescribed this argument.

30. Sidgwick, *The Methods of Ethics* (London: Macmillan, 1907; Indianapolis: Hackett, 1981), concluding chapter.

if his only aim was to escape to Brazil? And would it be rational for him to escape, even if his only aim was to surrender?[31]

Kagan presents other objections to the argument I have been discussing. I called S implausibly "intolerant" in its claims about the rationality of different aims. Kagan writes that, "to the extent that S can be convicted" of such intolerance, so can the Present-aim Theory (P). Theory S gives supremacy to one particular aim. But P is equally intolerant, he claims, since P does the same (Kagan, pp. 750–51).

Theory P does, in one sense, give supremacy to one aim. Like any other theory about rationality, P answers the question, What do we have reasons to try to achieve? In answering this question, P gives us a single *ultimate aim,* or what Kagan calls a "master function." On the simplest version of P, our ultimate aim should be the achievement of our present aims, *whatever* these are.[32] Is this version of P intolerant? It does not claim *any* aim to be less rational than any other. Nor does it insist that we be guided by any particular aim. "Our present aims, whatever these are" is not a *restrictive* master function. But S's master function is restrictive. We often have aims which conflict with the pursuit of our own long-term self-interest. Theory S claims that, in all such cases, it is irrational to try to achieve these aims.

Kagan writes that, if we turn from master functions to ordinary aims, S is again as tolerant as P, since S need not claim any of these aims to be "rationally unacceptable or inferior" (Kagan, p. 751). But, if these other aims are not inferior, why should self-interest be the master function? If these other aims are no less rational, why should we believe that, when they conflict with our own self-interest, it is irrational to act upon them? (Kagan might object that, in deciding between S and P, we can ignore the rationality of *aims.* He suggests later that our only question is about the rationality of *acts* [Kagan, p. 756]. But, as I argued [pp. 132–33], this is a weak reply.)

In another comment on my argument, Kagan writes, "The intuitions appealed to are not ones that need to trouble the Self-interest Theorist" (Kagan, p. 750). As he later writes, "Even if we can construct cases where our intuitions strongly support P over S, it seems likely that we can also construct cases where our intuitions support S over P" (Kagan, p. 757). I believe that this is not so, if we are considering what I called the "Critical" version of the Present-aim Theory, or CP.[33] This theory can

31. Kagan accepts these claims only if they are taken in a tenseless way. On his view, this criminal has a reason to surrender if this was *once* his aim, or will *later* be his aim. Whether it is rational for him to surrender does not depend, in any special way, on whether this is *now* his aim. I discuss this view further in Appendix G below.

32. This is the uncritical version of P which I called the "Instrumental Theory" (pp. 117–20).

33. According to CP, we have reason to fulfill our present aims if and only if these aims are rational. As I argued, this is the best version of the Present-aim Theory (pp. 117–26, 131–36, 188–95).

claim both that a rational agent should care about his own self-interest and that this concern should be temporally neutral.[34] I believe that this is as far as most people's intuitions go. They do not support the claim that it is irrational to care as much about anything else. Only if we add this claim does CP coincide with S.[35]

Kagan makes a stronger claim about our intuitions. After writing that one of my later arguments shows S to be "counter-intuitive," he suggests that a Self-interest Theorist should "simply dismiss such intuitions as incorrect." This argument's "fundamental flaw," he claims, is that "in the absence of an account of the basis of a theory of rationality, rival theories can only be played off against [our] intuitions . . . and such intuitions can never be decisive" (Kagan, p. 757). Like Schultz, Kagan here assumes that our choice between rival theories cannot depend on their plausibility—that only decisive arguments make any difference. But even in the physical sciences our choice between some theories must depend on their plausibility. And Kagan admits that my argument shows his version of S to be "counter-intuitive" (Kagan, p. 757).

WHAT WE TOGETHER DO

As Bart Gruzalski shows, my third chapter needs to be revised (Bart Gruzalski, "Parfit's Impact on Utilitarianism," in this issue). I described cases where

1. by acting in certain ways, one group of people harm another group,

though

2. the effects of each act are so widely spread that they are either trivial or imperceptible.

Thus my imagined Harmless Torturers inflict great suffering on their victims, even though what each torturer does makes no victim's pain perceptibly worse.[36]

34. This claim does not conflict with my p. 314.

35. Theories S and CP coincide, I wrote, if CP claims that the desire to pursue self-interest is "rationally required to be our strongest desire" (pp. 131–33, 194). Kagan objects that S might be indirectly self-defeating: it might be worse for someone if he always tried to pursue his own self-interest (Kagan, p. 758, n. 17). As I wrote, S would tell such a person that, though his own self-interest should be his one ultimate aim, it should not always be his strongest desire (pp. 5–10). In my claims about CP and S I deliberately ignored this complication (pp. 128–29). But if these claims are revised, by changing "strongest desire" to "ultimate aim," Kagan's objection is met. Kagan also objects that, even if a version of CP coincides with S, it will still claim "that only *present* desires can directly generate reasons. . . . But the Self-interest Theory *disagrees* about this fundamental point" (Kagan, p. 759). As I explained (pp. 121–22), there is no disagreement here. (In this explanation, to meet Kagan's other objection, "desire" should be changed to "aim.")

36. My discussion of these cases derives from B. Barry, *Political Argument* (London: Routledge & Kegan Paul, 1965), pp. 328–30; and J. Glover, "It Makes No Difference Whether or Not I Do It," *Proceedings of the Aristotelian Society*, vol. 49, suppl. (1975).

Why are these torturers acting wrongly? One explanation appeals to the effects of what they together do. Another explanation claims that there are imperceptible harms and benefits. Because this claim is controversial, I wrote, "It is better to appeal to what groups together do" (p. 82).

This appeal corrects what I called the "Second Mistake": that of ignoring the effects of sets of acts (p. 70). It also avoids some of the problems raised by what I called the "Fifth Mistake," since we can regard this as "a special case of the Second Mistake." But, as Gruzalski shows, my statement of the Fifth Mistake (p. 75) must be revised. It can become

> M5: An act cannot be right or wrong because of its effects, if the effects of this particular act are imperceptible.

This is a special case of

> M2: If an act is right or wrong because of its effects, the only relevant effects are the effects of this particular act.

Since M2 is false, so is M5. An act *can* be right or wrong because of its effects, even if the effects of this particular act are imperceptible. These are not the only relevant effects. The act may be right or wrong because it is one of a set of acts which *together* have perceptible effects. In my imagined case, what each torturer does makes no perceptible difference, but what they together do makes a great difference.[37]

Besides misstating the Fifth Mistake, I misstated my appeal to what groups together do. I intended to avoid the claim that there are imperceptible harms and benefits; but, unless we accept this claim, the principle to which I appealed cannot explain why the torturers are acting wrongly. I should have scrapped this principle—C12—and referred back to

> C7: Even if an act harms no one, this act may be wrong because it is one of a set of acts that together harm other people. [P. 70]

This explains more simply what I was trying to explain.[38]

37. Suppose we believe both that there are imperceptible harms and benefits and that an act like that of my Single Torturer (p. 82) is wrong because of its effects. On this view, even if the effects of an act are imperceptible, *these* effects *may* make this act wrong. It would then not be true that M5 is merely a special case of M2. This is why I failed to state M5 so that this could be true.

38. Similar remarks apply to cases where some group together benefits another group. In discussing the Drops of Water (pp. 76–78) I should have scrapped C10 and referred back to C7. There is another way in which I was confused. At the end of chap. 2 I wrote, "It is often claimed that, in those Contributor's Dilemmas that involve very many people, what each person does would make no difference. If this claim was true, a rational altruist would not contribute. But, as I argue in the next chapter, this claim is false." This is *not* what chap. 3 argues. Besides the two mistakes described above, I even misstated chap. 3's main aim. This was to show that, even when each act would make no perceptible difference, rational altruists would still contribute. (In some cases, though each act does not even make an *imperceptible* difference, we should appeal to what groups together do. This can be true, e.g., when effects are overdetermined, as in the case of my two murderers [p. 70].)

After showing the need for these revisions, Gruzalski writes, "Parfit . . . has not solved the original problem" (Gruzalski, p. 781). But I was discussing three questions: (1) What is wrong with the Fifth Mistake? (2) Why are the torturers acting wrongly? (3) When we appeal to what groups together do, whom should we count as members of these groups? With the revisions mentioned above, what I wrote answers 1 and 2.

In cases which involve imperceptible effects, 3 may be difficult. We cannot simply count, as members of the relevant groups, all those who act in certain kinds of ways. In some cases, the members of the relevant groups do not all act in the same way; in other cases, those who act in the same way are not all members of these groups. We may have to define these groups as those whose acts have certain imperceptible effects. We must then say which effects these are. A tempting answer is "imperceptible harms and benefits"; hence my claim that there are such things. But this claim raises what is called the Sorites Problem.

Here is Dummett's version of this problem. It is natural to assume that, if two painted cards look in every light the same to all observers, these cards must be painted with the same colors. But this assumption seems to imply that black is white. There can be a row of cards of which the first is white, and the last black, even though each card in the row looks the same as both of its neighbors. On this assumption, the second card must be white, so must the third, the fourth, and so on.

As this example shows, the Sorites Problem has nothing particular to do with ethics; it may arise whenever we consider imperceptible differences. And this problem does not yet have an agreed solution.[39] Gruzalski is right to claim that, since my chapter 3 did not solve this problem, it did not fully answer one of the questions that I raised. In my appeal to the effects of what groups together do, I did not fully explain whom we should count as members of these groups. This gap needs to be filled. There are many actual cases where, because the effects of each act are imperceptible, it is hard to tell who are the members of the groups who together harm or benefit other people.

Often, however, we *can* tell. Thus it may be clear who are the people who together cause pollution, or congestion, or soil erosion, or the depletion of many kinds of resources. And these people cannot defend their acts by appealing to a Sorites Argument. Such arguments cannot show that black is white; nor can they excuse the Harmless Torturers.

There is another point which I did not sufficiently explain. Gruzalski claims that, in my examples, "the rightness or wrongness of an action is a function of the consequences of others doing the same kind of action" (Gruzalski, p. 781). This could be taken to refer either to the *actual* consequences which are produced by some group who act in some way,

39. See M. Dummett, "Wang's Paradox," *Synthese* 30 (1975); C. Peacocke, "Are Vague Predicates Incoherent?" *Synthese* 46 (1981); and G. Forbes, "Thisness and Vagueness," *Synthese* 54 (1983).

or to the *hypothetical* consequences which would be produced if some group—or everyone—were to act in this way. These two interpretations give us very different moral views. The first is a Consequentialist view about the effects of sets of acts: a view which appeals to what will in fact happen. The second is the Kantian view expressed in the question, "What if everyone did that?" Kantians reject the answer, "No one else *will* do that," since their view does not appeal to what will in fact happen.

Gruzalski suggests that, in my appeal to what groups together do, I was moving from the Consequentialist to the Kantian view.[40] This is not so. I claimed that, because we *can* appeal to the Consequentialist view, we need *not* appeal to the Kantian view.

Though Gruzalski shows that my discussion needs to be revised, he does not undermine this claim. And it applies to many actual cases. It is therefore not enough to ask, "Will my act harm other people?" Even if the answer is no, this act may still be wrong because of its effects. It may be one of a set of acts that, like the acts of my imagined torturers, greatly harm other people.[41]

IS COMMON-SENSE MORALITY SELF-DEFEATING?

I argued that, in countless cases, Common-Sense Morality is self-defeating, and therefore needs to be revised. Arthur Kuflik challenges this argument (Arthur Kuflik, "A Defense of Common-Sense Morality," in this issue).

My main claims were these. According to Common-Sense Morality, which I called M, we have special duties to those whom I called "our M-related people." These include our children, parents, pupils, patients,

40. Thus he calls this appeal "a shift away from act consequentialism," "Kantian," and "Harrod's solution" (Gruzalski, pp. 779, 781). (But he also calls it "a multi-act version of utilitarianism" [Gruzalski, p. 781].)

41. Gruzalski suggests a different explanation of why the torturers act wrongly. "In such cases, we may analyze the consequences of *each* action in terms of threshold effects in the context of the behaviour of others. . . . On the threshold analysis those acts which fall below the pain causing threshold have imperceptible effects, whereas those at the threshold do not" (Gruzalski, p. 780; my italics). These remarks do not apply to the kind of case which I discussed. Consider the Pain-Reaction Test. At the start of an experiment, I am in mild pain. I am asked to say, when a bell rings every ten seconds, whether since the last ring my pain seemed to get worse. The following might happen. During each of these ten-second periods, the tester slightly increases some painful stimulus. Because the increase is so slight, it seems to me, after each period, that my pain did not get worse during this period. But after several minutes I must admit that my mild pain has become severe. There will of course be a ten-second period in or after which it seems to me, for the first time, that my pain is worse than it was at the start. This is the period in which, compared with its badness at the start, my pain will pass this noticeable threshold. But it will also seem to me that this pain did not get worse in this ten-second period. What the tester did during this period, though it caused this threshold to be passed, did not itself make any perceptible difference. Dummett cites this analogy: looking at a watch to see when the minute hand moves. Though this hand never seems to move, we can soon see that it must have moved. (For some other comments on Gruzalski's paper, see my Appendices C and D below.)

clients, fellow workers, and those whom we represent. There are countless cases where, if we all give priority to the interests of our M-related people, this will be worse for all these people. In some of these cases M tells us to give priority to these people. On my definition, M is here directly collectively self-defeating. Those who accept M should therefore move to a revised version R. According to R, what we should all ideally do, in these cases, is to give *no* priority to our M-related people. And this is what each of us should do if many others do so too (pp. 53–56, 95–108).

Kuflik presents many objections to these claims. He argues that they apply only to a "very narrow" range of cases, that only an "amoral" view could be in my sense self-defeating, and that to be in this sense self-defeating is not a defect in a moral view. I explain in Appendix E below why I disagree.

Some of Kuflik's other objections have more force. My main examples I called "Parent's Dilemmas." In these cases

i: each of us can either (E) benefit his own children or (A) give a greater benefit to the children of others, and
ii: because we cannot communicate, or for some other reason, what each of us does will not affect what the others do.

Each of us would know that, *whatever* others do, it will be better for his own children if he himself does E rather than A. But if we all do E rather than A this will be worse for all our children.

Many Parent's Dilemmas involve a public good: an outcome which would benefit our children whether or not we help to produce it. Some examples are better schools and local government, clean air and water, law and order, and the preservation of many resources. Each of us can either (A) contribute to this public good or (E) fail to contribute and spend what he saves—whether in money, time, or energy—directly on his own children.

According to R—my proposed revision of Common-Sense Morality—each of us ought to contribute if he believes that many others will do so too. To support this claim I wrote: "If any parent does not contribute when others do, his children will be *free-riders*. They will benefit from this public good at the expense of the children of contributors. They will benefit at their expense because (a) they will be benefited more than the children of contributors, and (b) this is true because each contributor did what was worse for his own children" (p. 101).

Kuflik objects that these claims do not *revise* Common-Sense Morality, since they are *already part* of this morality. If I am the only parent who does not contribute, most of us would think this unfair. Kuflik then argues that, on Common-Sense Morality, these cases are *coordination problems:* what each of us ought to do depends upon what others do. My failure to contribute would be wrong if I knew that every other parent

will contribute, but it would not be wrong if I knew that I would be the only contributor (Kuflik, pp. 790–92). I accept this objection. And I agree with Kuflik that it undermines one of my main claims.[42]

It will be easier to explain what this objection shows if I first discuss another similar objection. On my proposed revision R, we should all ideally contribute to the production of these public goods. Kuflik objects that this claim is also part of Common-Sense Morality. He concludes that, if we knew that we all knew that we were all conscientious M-believers, we could always solve these coordination problems. Even if we could not communicate, each of us would know that everyone else will give no priority to his own children. Since each could rely on others, this is how we would all act (Kuflik, pp. 792, 796).

I doubt these claims. Suppose that you and some stranger cannot communicate. Each of you knows that each could either save his own children from some harm, or save the other's children from another greater harm. If you knew that this stranger was an M-believer, would you assume that he believes that, ideally, both of you should give no priority to protecting your own children? I suspect that you would not be sure of this. Nor would you be given this assurance if you knew this stranger to be Dr. Johnson, or Sir David Ross, or some other notable exponent of Common-Sense Morality. One ground for doubt is that many M-believers have never considered Parent's Dilemmas. (This is not because there are few such cases. Many people have never considered the very common similar cases now called Prisoner's Dilemmas.)[43]

Kuflik makes another claim which may seem to show that, in Parent's Dilemmas, conscientious M-believers would give no priority to their own children. Return to the case where each can either (E) benefit his own children or (A) give a greater benefit to the children of others. Kuflik writes that, if we all do A, this is "the best that each can do for his own child without being unfair or unjust to the other child. . . . When each benefits the other's child in the expectation that the other will benefit his, each does the best he can for his own child . . . within the moral limits which M lays down" (Kuflik, pp. 793, 794). If this claim was true, conscientious M-believers *could* be relied upon to benefit each other's children. Each would surely want, within the limits which M lays down, to do the best he can for his own children.

This claim is not true. If we all do A, benefiting the children of others, each does *nothing* for his own children. If instead we all do E, giving the lesser benefits to our own children, what we are doing would

42. As Kuflik writes, my first (confused) reaction was to accept this objection only in two-person cases.

43. Kuflik discusses Prisoner's Dilemmas in his appendix (Kuflik, pp. 802–3). That these cases are common I claimed in my sec. 22; their relation to Parent's Dilemmas I discussed in sec. 36.

not be unfair. *This* is how, within the limits which M lays down, each can do the best he can for his own children.[44]

It may be objected that, as Kuflik points out, we would "strongly prefer" that we all do A rather than E, since this would be better for all our children (Kuflik, p. 795). But, though this is true, it does not show that we would all do A. Suppose that each of us believes that others will do E. We would believe this, for example, if this is how most parents have acted in the past. Each of us would then be likely to do E however strongly he prefers that we all do A. As Kuflik writes, each would be specially anxious not to be the only parent who does A. If any parent does A when others do E, what this parent does is not only worse for his own children but also unfair to these children.

I can now explain what I believe that Kuflik shows. I argued that, because Common-Sense Morality is in these cases self-defeating, M-believers ought to accept my proposed revision R. On this revision,

> Ri: each of us should do A if he believes that many others will do A,

and

> Rii: it would be morally best if we all did A.

Kuflik claims that, because Ri is already part of M, these cases are for M-believers mere coordination problems. And he claims that, because Rii is already part of M, a community of M-believers could always solve these problems.

I have questioned the second of these claims. But I accept the first, and this is enough to undermine part of my conclusion. M is directly collectively self-defeating if

1. it is certain that, if all rather than none of us do what M claims that we should do, we will make the outcome worse for all our children, or
2. we will make the outcome better for all our children only if we all do what M claims to be wrong.[45]

I assumed that, on Common-Sense Morality, we should all do E, the act which benefits our own children. I assumed that, if we all did A, giving greater benefits to the children of others, we would all be doing what M claims to be wrong. On these assumptions, 1 and 2 are true, and M is here directly self-defeating. But, as Kuflik points out, M does *not* tell us

44. Kuflik's claim would be true if we both communicate and make a joint conditional promise that we shall all do A. As I argued, making this promise is the best that each can do for his own children (p. 107); and M would tell us to keep this promise. But if we make no such promise, and do not in other ways affect what most other people do, Kuflik's claim is not true.

45. These claims apply the definition on my p. 55, if we assume that, in these cases, the M-given aim of each is that the outcome be better for his own children. (This is not the assumption that, according to M, each of us should *always* give this aim priority.)

all to do E. On Common-Sense Morality, we would not be acting wrongly if we either all do E, or all do A. Since 1 and 2 are not true, M is *not* directly self-defeating.

Kuflik undermines this claim; but he accepts another part of my conclusion. Though he denies that R revises M, he agrees that M-believers ought to believe R. And this is the more important part of my conclusion.[46]

My mistake came in my description of what people actually believe. I claimed that, in Parent's Dilemmas, most people believe that they ought to give priority to their own children. Kuflik points out that this is not true. The moral view which I called M is *not* Common-Sense Morality. I must therefore revise my argument, so that it takes a conditional form. It should claim that, *if* people accepted M, their moral view *would be* directly self-defeating.

My mistake makes me cautious. But I suggest that, in this conditional form, my argument still shows that M-believers ought to believe R. Common-Sense Morality ought to include R because, if it did not, it would be directly self-defeating. Kuflik's claims do not, I suggest, undermine this argument.[47]

I gave another argument for the same conclusion (pp. 107–8). Suppose that, in a Parent's Dilemma, we can all communicate. If we are all conscientious M-believers, each ought then to tell the others that he promises to do A, provided that everyone else makes the same promise. As I argued, making this joint conditional promise is the best that each can do for his own children. This is how each can ensure that the others will give to his children the greater benefits (pp. 107–8).

As Kuflik claims, in most Parent's Dilemmas communication would be possible. But it may be difficult, or costly. This is especially likely if, as is now often true, the case involves thousands or millions of people. The remedy is a general promise, covering all Parent's Dilemmas. This promise should also cover all of our other special obligations to our other M-related people—such as our parents, colleagues, pupils, patients, clients, and constituents. When M-believers reach maturity, they should all pledge to one another that, in the countless cases that I described, they will all give no priority to the interests of their M-related people. This pledge would make R a part of M.[48]

I conclude that, as Kuflik would agree, R ought to be part of Common-Sense Morality. In these countless cases we should not believe that we

46. If M-believers ought to believe R, little turns on the question whether they already do. And there are two ways in which this question is unclear. Common-Sense Morality is not a single, well-defined view but merely what is in common to the various views which are most widely held. Nor is there a sharp distinction between what people do believe and what, if they considered certain questions, they would believe.

47. As I explain in Appendix F below this argument could also be expressed in a milder way.

48. There are several practical obstacles to this joint promise, but these do not affect this ground for claiming that R ought to be part of M. If we should all ideally promise to do what R tells us to do, this is what we should all ideally do.

should all carry out our special obligations. Acting in this way is at most a defensive second-best.[49]

FUTURE GENERATIONS

James Woodward questions my discussion of the Non-Identity Problem (James Woodward, "The Non-Identity Problem," in this issue). My main claims were these. Certain choices may predictably cause some future people to be killed or injured, or to be badly off in other ways. These seem to be bad effects, which give us moral reasons not to make these choices. But it may also be predictable that, if we had *not* made these choices, these particular future people would never have existed. We may therefore know that, if we make these choices, this will not be worse for these future people. Does this remove our moral reasons not to make these choices? I claimed that it does not. Are these reasons just as strong as they would be if these choices *would* be worse for these future people? I claimed that they are. What are these reasons? How should they be explained? I claimed that these reasons cannot be fully explained either by appealing to people's interests, or by appealing to people's rights. We therefore need a new theory about beneficence. Apart from the principle which I called Q, I failed to discover this theory (pp. 351–79).[50]

Woodward argues that I underestimated what we can explain by appealing to people's interests and to people's rights. We have, he suggests, little need for a new theory. Some of Woodward's claims I find plausible. But, since I am not sure how these claims are connected, part of what follows is an attempt to bring these claims together, and to state them in a clearer way.

Consider what I called the Risky Policy, which uses nuclear energy. We know that, if we choose this policy, this may in the distant future cause a catastrophe, in which released radiation kills or injures thousands of people. Following Woodward, I shall call these "the nuclear people." As I argued, we would also know that, if instead we choose the Safe Policy, it will be *different* people who will later live, and be spared the catastrophe (pp. 351–61, 371–73).

What is the objection to our choice of the Risky Policy? Woodward suggests that we should compare

> the situation of the nuclear people under the nuclear policy (when they are killed, injured, etc.) and an (unattainable) baseline situation in which the nuclear people exist and these violations of their rights do not occur. . . . On [this] approach . . . we resist the temptation to think just in terms of . . . how well off overall a person is. . . .

49. It is at most what each of us should do when he believes that this is what many others have done, or will do (pp. 99–102).

50. In this chapter of my book, and in what follows, I assume both that, in the different outcomes, the same number of people would exist, and that, in causing someone to exist, we cannot thereby benefit this person. My later chapters dropped these assumptions.

> We thus find it natural to think of the choice of the Risky Policy as harming the nuclear people . . . or even as worse for them (worse with reference to the above baseline) even though the overall effect of that policy is to leave the nuclear people no worse off than they would be under any possible alternative policy. [Woodward, pp. 817–18]

If this view is defensible, it would provide a complete solution to the Non-Identity Problem. And it would have wider significance.[51]

Woodward claims that the relevant comparison is with the *unattainable* alternative in which the nuclear people both exist and are not injured. To assess this claim, we must know in other cases what the relevant comparison would be. Suppose that, in a crash, you have become unconscious. A surgeon amputates your arm because this is the only way to save your life. What this surgeon does is better for you than every possible alternative. But, on a view like Woodward's, he might be acting wrongly. The relevant comparison might be with the impossible alternative in which he saves both your life and your arm.

How should we state Woodward's view so that it condemns the Risky Policy, but does not condemn this surgeon's act? In the case of the Risky Policy, and the other cases of this kind,

1. an act causes someone to be badly off, but
2. this act is not worse for this person than any possible alternative, because
3. if this act had not been performed, this person would never have existed.

On Woodward's view, 1 makes such an act wrong. Statement 2 should be ignored because what makes it true is 3. In the case of the surgeon's act, 1 is true: this act causes someone to lose an arm. Statement 2 is also true, because

4. if this act had not been performed, this person would have died.

If Woodward's view is not to condemn the surgeon's act, he must claim that in this case 2 should *not* be ignored. But what is the difference? Why should 2 be ignored when it depends on 3 but not when it depends on 4?

We might say, "Statement 4 has a moral significance which 3 lacks. It matters morally that, if the surgeon had acted differently, you would have died. It is morally irrelevant that, if we had not chosen the Risky Policy, the nuclear people would never have existed."

I believe that, for *practical* purposes, this fact about the nuclear people is indeed irrelevant. On my view, the Non-Identity Problem never affects what we ought or ought not to do. There is an objection to any

51. It is often claimed (e.g., on my p. 69) that we do not objectionably harm someone if we know that our act will not be worse for this person than any possible alternative.

act which causes a future person to be badly off, even if this act will not be worse for this person than any possible alternative. And this objection is *as strong* as it would be if we imagine away the Non-Identity Problem: if we suppose that, because 3 is false, this act *will* be worse for this person. These claims express what I called "the No Difference View" (pp. 366–71).

Woodward makes a further claim. If we imagine away the Non-Identity Problem, our choice of the Risky Policy would, he writes, be "just as wrong (*and for the same reasons*)" (Woodward, p. 817; my italics). On this view, the Non-Identity Problem does not even make a *theoretical* difference. It does not weaken the objection to our choice because it does not affect what this objection is.

Woodward's objection to our choice is that it harms the nuclear people and violates their rights. On his view, it is irrelevant that this choice will not be worse for these people than any possible alternative would have been. If this is so, I asked, how can we avoid condemning the surgeon's act? Why is it relevant that this act is not worse for you than any possible alternative?

Someone might suggest, "In the special cases which raise the Non-Identity Problem, whether an act objectionably harms a person does not depend on what alternatives were possible. In these cases the relevant comparison is with an unattainable baseline. In other cases this is not so." But this suggestion explains nothing. Why is there this difference?

Woodward suggests a better answer. According to what he calls the "no worse off argument," an act cannot objectionably harm a person, or violate this person's rights, if this act is not worse for this person than any possible alternative (Woodward, pp. 808–9). On the suggestion just mentioned, we should reject this argument only in the special cases which raise the Non-Identity Problem. Woodward rejects this argument in a much wider range of cases. He presents two main objections. (In what follows "worse" will mean "worse than any possible alternative.")

Woodward claims that many of our rights "protect highly specific interests," such as our interest in having promises kept or avoiding bodily injury. If an act "violates" one of our specific interests, the fact that this act will not be worse for us "is not always an adequate response" (Woodward, p. 809). In some cases this fact is not enough to excuse or justify such acts.

Which are these cases? When is this fact not enough to justify such acts, and when *is* it enough? Woodward's answer seems to appeal to his other main objection to the "no worse off argument." This is provided by the "non-consequentialist [idea] . . . that actions and not just states of affairs are distinctive objects of moral assessment." It makes a difference *how* and *why* acts have their effects. Thus the direct and intended effects of an act matter more, in its assessment, than its remote and unintended effects (Woodward, p. 809).

Woodward's claims could be combined in

> THESIS *T*: Suppose (1) that an act predictably violates one of a person's "specific interests," though (2) the agent knows that, because this act will have some other effect, it will not be worse for this person. Clause 2 provides an excuse only if (3) this other effect is both intended and directly caused.

If this view is defensible, it would explain why the surgeon does not act wrongly. When he amputates your arm, he violates your specific interest in avoiding bodily injury. But his act is not worse for you because it also saves your life; and he both intends and directly causes this effect. No such claim applies to our choice of the Risky Policy. This is not worse for the nuclear people because, as I wrote, they "owe their existence" to this choice. Woodward protests against this use of "owe." Since we do not directly cause the nuclear people to exist, we can claim no credit for their existence. Nor is their existence any part of what we intend.

Though Woodward does not explicitly assert Thesis T, something like it seems to be assumed in much of his discussion, and it fits most of his examples. Thus T implies that the Nazis wrong Frankl, even though he ends up better off than he would otherwise have been. The effect which is good for him—the wisdom that his suffering brings—was neither intended by the Nazis, nor directly caused by what they did. And T condemns the airline which denies Smith a ticket because he is black. The effect which is good for Smith—his absence from the plane which crashed—may perhaps be directly caused by this act. But the airline did not intend to save Smith's life. Thesis T seems to explain why, in such cases, we should reject the "no worse off argument."[52]

But consider

> *The Two Women.*—While Carla is pregnant she learns that, unless she takes some treatment, there is a risk that her child may have a certain handicap. She decides not to take this treatment. As a result her child, Carl, is handicapped.
>
> While Paula is trying to become pregnant, she learns that, if she conceives a child now, there is a risk that it may have the same handicap. If she waits two months before conceiving a child, there would be no such risk. She decides not to wait. As a result her child, Paul, is handicapped.

Carla's act is worse for Carl. If she had taken the treatment, Carl would not have been handicapped. But Paula's act is not worse for Paul. She

52. At one point Woodward seems to reject T. He suggests that the "no worse off argument" remains "problematic" even if we assume that the agent's aim is to benefit the affected person. Thus the objection to the Nazis would not, he writes, be "fundamentally altered" even if their aim had been to benefit Frankl (Woodward, p. 810). But it surely makes a difference that, when the surgeon amputates your arm, his aim is to save your life. (Woodward may be misled by his example. If we try to imagine that the Nazis' aim was to benefit their victims, the example becomes unclear.)

could not possibly have had Paul without giving him this handicap. If she had waited, the child she would have had later would have been a different child.

There are clearly objections to what these women do. But are these the same objection? If the objections are different, is one of them stronger?

On the view presented in my book, these objections are equally strong. This suggests that there is the same objection to each act. Since Paula's act is not worse for Paul, this objection cannot appeal to Paul's interests. It must appeal, I argued, to a new theory about beneficence (pp. 366–71).

When he discusses a similar case, Woodward writes, "Consider conceiving a child which one knows will be genetically handicapped. . . . I assume the case to be one in which it would be impossible to have *that* child without the handicap. . . . I think that it is difficult to frame a plausible theory of rights according to which conceiving a genetically handicapped child (whose life is worth living) violates his rights. If this is correct . . . there is no rights-based objection to conceiving such a child" (Woodward, n. 12). On this view, Paula does not violate Paul's rights. But this view does not apply to Carla. She could have had Carl without giving him this handicap; and, if we have any rights, we have a right not to be avoidably handicapped.

Why does Woodward claim that there is no rights-based objection to an act like Paula's? In the earlier passage quoted above, he suggests that we should ignore the fact that, if we had not chosen the Risky Policy, the nuclear people would never have existed. The relevant comparison is with the "unattainable baseline" in which these people would have both existed and not been injured. On this view, our choice violates these people's rights because it is worse for them than this impossible alternative. If this view is defensible, we could also ignore the fact that, if Paula had acted otherwise, Paul would never have existed. Paula's act could violate Paul's rights because it is worse for him than the impossible alternative in which he would have both existed and not been handicapped. Given the similarity between these cases, why does Woodward claim that an act like Paula's does not violate her child's rights?

He may be assuming what I called Thesis T. On this Thesis there is an objection to an act if (1) it predictably violates one of a person's specific interests. This objection may be answered if (2) the agent knows that, because his act will have some other effect, it will not be worse for this person than any possible alternative. But 2 answers this objection only if (3) the agent both intends and directly causes this other effect. Thus, in the case of the Risky Policy, 1 and 2 are both true. Though our choice of this policy will lead to the injuring of the nuclear people, it will not be worse for these people because it will also cause their existence. But we neither intend nor directly cause this effect. On Thesis T, our choice is wrong because claim 3 is not true.

In the case of Paula, however, all three claims are true. By refusing treatment, Paula predictably causes her child to be handicapped. This violates one of this child's specific interests. But Paula also knows that her act will not be worse for this child than any possible alternative. And this is true because of an effect, the existence of this child, which Paula both intends and directly causes.

In considering this case, Woodward has two alternatives. He might keep his claim that there is no rights-based objection to such an act. This would allow him to appeal to Thesis T. If instead he withdraws this claim, he would need to revise T. I shall consider both alternatives.

Suppose first that he keeps his claim. As he says (Woodward, n. 12), if the objection to Paula's act cannot appeal to her child's rights, it must appeal to a principle like Q, which compares different possible lives. The objection must be that Paula's act makes the outcome worse in the way described by Q.[53] Carla's act, in contrast, violates her child's right not to be avoidably handicapped. On this version of Woodward's view, there are different objections to these two acts. And the objection to Carla's act must be stronger. Violating such a right must be more serious than making the outcome worse in the way described by Q. Those who appeal to rights all give them some priority over what Woodward calls "consequentialist considerations."

Remember now what I called the No Difference View. On that view, though the Non-Identity Problem raises theoretical questions, it never affects what we ought or ought not to do. There is an objection to any act which causes a future person to be badly off, even if this act will not be worse for this person. And this objection is as strong as it would have been if this act *had* been worse for this person. This was my view.

On the account just given, Woodward must reject this view. On this account, only Carla violates her child's rights. Paula's act is less wrong because it is wrong for a different and less weighty reason. If this is so, I should not have claimed that the Non-Identity Problem has no practical importance. Since it sometimes undermines an appeal to people's rights, it *can* affect what we ought to do. Thus, in the similar case that I discussed, it would affect which of my medical programs we should choose (pp. 367–69). On this version of Woodward's view, pregnancy testing would prevent the violation of children's rights, while pre-conception testing would merely make the outcome better in the way described by Q. Since rights have priority, pregnancy testing would be the better program. And it could be better even if pre-conception testing would do more to reduce the incidence of this handicap. Similar claims apply to many other cases.

53. According to Q, if in either of two outcomes the same number of people would ever live, it would be [worse] if those who live are worse off, or have a lower quality of life, than those who would have lived.

Though these conclusions seem to be implied by what Woodward writes, it is not clear that he would accept them. "We need *not* agree," he writes, "that the no worse off argument ever defeats a claim of rights violation" (Woodward, p. 815). If he believes that this argument *never* defeats such a claim, this suggests that he accepts the No Difference View.

If he does accept this view, he must withdraw his claim that there is no rights-based objection to the conception of a handicapped child. He could then appeal only to some revised version of Thesis T.[54] Rather than considering how T ought to be revised, I shall end by comparing Woodward's view with mine.

In the cases that we are discussing, certain choices cause some future people to be badly off; but, because these people would not have existed if we had chosen otherwise, these choices are not worse for these people. On the No Difference View, this fact has no practical importance, since it does not weaken the objections to these choices. But there is a theoretical problem: these objections must be explained. They cannot be fully explained, I argued, either by appealing to people's interests or by appealing to people's rights. In some cases we must appeal to the new theory about beneficence for which I vainly searched (pp. 361–66).

Woodward shows that I underestimated what we can explain by appealing to people's rights.[55] But he accepts my claim that we need this

54. There are other cases in which T needs to be revised. Suppose that one of two surgeons amputates my arm so that the other can save my life. The first surgeon does not act wrongly, though he does not directly cause the saving of my life. We might revise Thesis T so that clause 3 refers to what the agent and others *together* do. But suppose next that I am unconscious after some accident. My arm is trapped in wreckage and is bound to be lost through frostbite. Because a surgeon knows this fact, he amputates this arm so that he can save someone else's life. He does not act wrongly, because he knows that his act will not be worse for me. But what makes this true—the effect of the frostbite—is not caused by him, either alone or with others. Thesis T must again be revised, so that it does not condemn this surgeon's act. It is unclear what the revision should be. This act may seem to be justified simply because 2 is true—simply because this surgeon knows that it will not be worse for me. But Woodward's main claim is that 2 does not, by itself, provide an excuse.

55. I assumed that, in these cases, an appeal to people's rights must take an unusual form. The objection must be that we should not cause people to exist whose rights cannot be fulfilled. But, as Woodward says, the objection can simply be that we should not cause people's rights not to be fulfilled. (Woodward also corrects my discussion of my imagined Fourteen-Year-Old Girl (pp. 358–61). This girl deliberately has a child, though she knows that, because she is so young, she will fail to give this child a good start in life. As Woodward claims, when this girl later fails to care properly for her child, she will be acting wrongly. In having a child when she is fourteen, she makes it more likely that she will later act wrongly in this way. This fact is enough to explain why she made the wrong decision (Woodward, pp. 815–16). (We cannot make similar claims about the other cases that I discussed. These cases also involve decisions which will cause future people to be badly off; but this will not be because these decisions make it more likely that we shall later act wrongly. Thus, if we choose the Risky Policy, it will not be an act but an earthquake which will later, by releasing radiation, injure the nuclear people.)

new theory. And he should still accept this claim even if he decides that the objection to Paula's act can appeal to her child's rights. As I argued, such an appeal cannot wholly explain this objection (pp. 375–76).

Here is a simpler version of my argument. Suppose that Paula knew that, if she had a child now, this child would be handicapped. She also knew that, if she waited, she would later have a different child who would not be handicapped. Though she knew these facts, she decided not to wait. Compare her with another woman, Petra. This woman knew that, even if she waited, any child of hers would have the same handicap. Though she knew this, Petra intentionally had a child, Peter. Suppose finally that this handicap, though not trivial, is not of a kind which makes these children's lives doubtfully worth living.

There is a moral difference between these cases. There may be an objection to what Petra does. But there must be a stronger objection to what Paula does. While Petra's only alternative was to remain childless, Paula's alternative was to wait and have a different child who would not be handicapped. Paula's decision not to wait must be worse than Petra's decision not to remain childless. And this difference cannot be explained by appealing only to the effects on these two children. What Petra does to Peter is the same as what Paula does to Paul. The stronger objection to Paula's act cannot appeal only to Paul's rights. It must also appeal to the fact that, unlike Petra, Paula could have had a different child who would not have had this handicap. It must appeal to something like my principle Q, which covers the different possible children whom Paula could have had.[56] As I argued, similar claims apply to many other cases, such as our choice between conserving or depleting resources (pp. 361–66). Even on this revised version of Woodward's view, my conclusion therefore stands. An appeal to rights cannot wholly solve the Non-Identity Problem.

Woodward ends with comments on my discussion of overpopulation. He suggests that I often sever the connection between such problems and any actual choices that human beings will ever face (Woodward, p. 829). But, though some of my imagined outcomes were not possible objects of choice, others were (pp. 382–87, 391–417, 438–41). Woodward then suggests that, if we merely ask about the relative badness of these outcomes, our question is hard to answer, and hard even to understand (Woodward, p. 830). I agree that this question can be hard to answer. For example, it may seem unclear whether the mere existence of extra people could make the outcome worse. But the difficulty here is with this particular comparison, not the nature of the question. Suppose that if we choose to conserve resources a certain number of people would later live; if instead we choose depletion, there would later be the same number of people, who would all be worse off. Most of us would firmly

56. It might be objected that Paula's act is worse for other people. But we could assume that this is not true.

believe that this second outcome would be worse. Consider next the outcomes which I called A and Z. In A there would be billions of people living, all of whom would have a very high quality of life. In Z there would be many more people, all of whose lives would be barely worth living. According to what I called the Repugnant Conclusion, if Z had enough extra people, it would be *better* than A. Most of us find this conclusion hard to accept. We believe that Z would be worse than A. And such beliefs are one of our chief reasons for concern with overpopulation.

These beliefs are surprisingly hard to defend. As I argued, it is hard to avoid the Repugnant Conclusion (pp. 387–406, 419–41).[57] Woodward suggests that this does not matter. On his view, we need not deny that Z would be better than A. It is enough to claim that, if we could produce one of these two outcomes, it would be wrong to produce Z. This would be wrong, Woodward suggests, because the inhabitants of Z would have unfulfilled rights to have the resources which would give them better lives (Woodward, p. 822).

This suggestion may fail. The inhabitants of Z, rather than being people who have only piglike pleasures, might *be* pigs. Since *these* inhabitants of Z could not have better lives, producing Z would not violate their rights. More important, such a suggestion could not adequately solve our problem. Much of our moral thinking rests on beliefs about the relative badness of different outcomes. If these beliefs are hard to defend, we cannot merely substitute beliefs about the morality of acts. Here is another example. On some versions of the Average Principle, an event makes the outcome better if it causes those who are living to be, on average, better off. On this view, it might make the outcome better if all except the best-off people dropped dead. Even if this catastrophe would be worse for all of the survivors, it might raise the average quality of life. We might say, "Though this catastrophe would make the outcome better, it would be wrong deliberately to bring it about." But this would not sufficiently defend this view.

CONCLUSIONS

I shall not try to summarize this long discussion; but it seems worth listing my conclusions. Wolf points out that, if we became Reductionists, this might have certain bad effects. Kagan notices a gap in my defense of the Present-aim Theory. Gruzalski shows that, since I did not solve the Sorites Problem, my discussion of imperceptible effects did not answer all of the questions that I raised. Kuflik corrects my description of Common-Sense Morality. And, in his account of rights, Woodward may have shown that the Non-Identity Problem has practical importance. There are, of course, many other ways in which my book needs to be revised.

57. I revise this argument in my "Overpopulation and the Quality of Life," in *Practical Ethics*, ed. P. Singer (Oxford: Clarendon Press, 1986).

APPENDIX A

THE CARTESIAN VIEW

Wolf claims that even Cartesians cannot easily justify our concern about our own future. "If the Cartesian Ego view is correct, then my interest in myself simply amounts to an interest in the Cartesian ego that is me. Thus insofar as I care about some future person solely on the grounds that he will be *me,* I care about him on the grounds that his consciousness and my present consciousness have an ego in common. But, given that a Cartesian ego is independent of personality, memories, even physical and psychological continuity, surely *that* is not a very strong or sensible ground for caring about someone" (Wolf, p. 707).

This ground seems to me sensible enough. Suppose I know that I shall later be in pain. I am then told that, before this happens, I shall be made to lose all my memories and have my character transformed. Do these facts remove my reason to fear the pain? Williams argues that it does not. On his view, I have as much reason to fear this pain whatever psychological changes precede it. A Cartesian could add "or physical changes." On the Cartesian view, it will be just as much *me* who is in pain. This is not, as Reductionists might say, merely like the truth that Germany is still just as much Germany. It is a deep truth, and—a Cartesian would claim—it gives me a strong reason to be concerned. Wolf denies this claim. That this pain will be mine is not, she suggests, a sensible ground for concern. She would presumably agree that, when this pain *is* mine, this *will* be a ground for concern. Does this ground have no significance when and because it is in the future? This would have to be because no fact about the future can be a ground for present concern. This the Cartesian could justifiably deny.

APPENDIX B

THE PROSPECT OF DIVISION

In arguing that personal identity is not what matters, I appealed to the imaginary case where I divide. I claimed that, if we are Reductionists, we should regard the prospect of division as about as good as ordinary survival. For some people, it would be better, for others worse, depending on the details of the case (pp. 253–73). Wolf argues that this imagined case cannot show that personal identity is not what matters. In a typical life, she claims, the prospect of division would be the prospect of something horrible—something far worse than ordinary survival (Wolf, p. 715). I am not persuaded by Wolf's arguments for this claim. (Why does she assume that one of the two resulting people would be unable to earn a living?) But even if this claim was true it would not justify Wolf's conclusion. There could indeed be cases in which division would be far worse than ordinary survival. This would be so, for example, if both of the two resulting people would soon die. But my argument appealed to a different case. I assumed that, after my division, each of the resulting people would have prospects that are as good as mine would be if I did not divide. In *this* case, would my division be as good as ordinary survival? I claimed that it would. My relation to each resulting person contains, I argued, everything that matters in ordinary survival. If this is so, as Wolf seems to agree, it is irrelevant that neither of the resulting people would *be* me. It is therefore irrational to regard personal identity as what matters. To block this argument Wolf would need to show that there cannot be such a case—

that division would itself ensure that the two resulting people would have prospects which are worse than mine. This I believe she has not shown, and could not show.

This case also helps to show what the relation is between Wolf's other arguments and mine. Wolf argues that, if we ceased to care about identity, this would have various bad effects. I believe that these effects would be outweighed by good effects. Even if I am wrong, my arguments are not, as Wolf claims, "undermined" by hers. A better description, which she considers and rejects, is that these arguments are "incommensurable." Whether it is rational to care about identity is simply a different question from the question whether caring about identity would have good effects.

APPENDIX C

THE NONEXISTENCE OF PERSONS

Schultz and Gruzalski mention Scheffler's claim that, if my argument about "depth" could justify the Utilitarian rejection of distributive principles, it would also undermine the Utilitarian view. Utilitarians would have to claim that there are not such entities as persons, merely "series of events." As Scheffler writes, they could then argue: "It is irrational to try to compensate for the fact that some event in the past had a painful duration by providing some event in the present with a pleasant duration . . . they are *different* events." Scheffler claims that this argument backfires, since it forgets the fact that (in Schultz's words) "any moral theory requires some coherent notion of a minimally unified moral agent, capable of putting the theory into effect." (Compare Darwall's claim that "something can only be subject to moral norms if it is capable of agency, and whatever that requires a series of events doesn't seem to be up to it" [for these two references see Gruzalski, nn. 16–17].)

The argument described above does not, in this way, undermine itself. It appeals to the claim that my luxury did not in advance compensate Backup for his poverty. This argument can admit that both I and Backup are moral agents.

(Gruzalski makes a similar claim about the principle of desert. He suggests that, in rejecting this principle, a Reductionist should argue that there are no persisting moral agents. On his view, I cannot now deserve to be punished for a crime which "I" earlier committed, since this crime was in fact committed by "a radically different entity." And Gruzalski suggests that this argument is stronger than the argument I gave, according to which only the further fact carries with it desert [Gruzalski, pp. 765–66]. But Gruzalski must admit that there is a sense in which I am the same entity as the I who earlier committed my crime. His argument would have to claim that what makes me the same persisting entity, or the relation between me now and myself then, does not carry with it desert or guilt. And this is the kind of argument I gave. [As G. Strawson has suggested, Reductionism may also support a more familiar argument which denies that we have Free Will].)

APPENDIX D

GRUZALSKI'S OTHER CLAIMS

1. I discussed the claim that on the Reductionist View, we have no reason to be specially concerned about our own futures (pp. 307–12). Gruzalski challenges

this claim. "We can," he writes, "distinguish in an impersonal way between felt and unfelt experiences now . . . and felt and unfelt experiences in the future." Though these distinctions "presuppose no continuing self," they justify each person's special concern about those future experiences that will be *felt* (Gruzalski, p. 769). I can indeed distinguish in an impersonal way between my *present* experiences and the experiences of others, since I can refer to the former as *these* experiences, or as what Gruzalski calls "pain *here*." But I cannot now apply this distinction to *future* experiences. How can I now pick out those future pains which, in Gruzalski's words, *will* have the quality of being "felt"? I can refer to these either as *my* future pains, or as those future pains which will be connected in various ways to my present experiences. But the first way of referring to these experiences presupposes a continuing self, and the second cites the Reductionist claim to which Gruzalski presents his view as an alternative. As far as I can see, he provides no third way of drawing this distinction.

2. I defended the well-known view that Consequentialism, or C, may be indirectly self-defeating. It might make the outcome worse if we were all direct Consequentialists, or pure do-gooders. It may therefore be true of someone that it would be better if he had other dispositions. If this is so, C would claim that it would be wrong for this person to cause himself to lose these dispositions. But, if he has these dispositions, they will sometimes cause him knowingly to make the outcome worse. Because he has dispositions which C claims that he ought to have, he will act in a way which C claims to be wrong (pp. 26–37). Gruzalski argues that there cannot be such cases (Gruzalski, pp. 772–77).

In the case I gave, Clare's love for her child causes her to benefit him rather than giving a greater benefit to a stranger. Clare knowingly makes the outcome worse, and she is therefore doing what, as a Consequentialist, she believes to be wrong. I claimed that, even if we assume Determinism, Clare's act is voluntary in the sense which is required for it to be wrong. (As I wrote, "voluntary" may not mean "free" in the sense required for Free Will. Determinists can believe that, though we never deserve to be punished, our acts may be voluntary and therefore wrong.)

Gruzalski objects that, if we assume Determinism, Clare's act is not voluntary. "If Clare could not do differently given her dispositions, then her act is neither right nor wrong" (Gruzalski, p. 774, also p. 776). He remarks that this objection may seem to imply "that no actions are voluntary if determinism is true." But he denies that this is so. An agent's act is voluntary, he writes, "if he or she would have done otherwise had he or she so chosen" (Gruzalski, p. 775). Gruzalski then claims that Clare's act is not, in this sense voluntary, since, like a kleptomaniac, she would have acted as she did "whatever she had chosen." This last claim is incorrect. As I wrote, Clare would have acted otherwise if she had chosen otherwise (pp. 32–33). On Gruzalski's own analysis, Clare's act *is* voluntary.

3. Following Adams, I claimed that Motive Utilitarianism is not just a special case of Act Utilitarianism (p. 505, n. 10). Gruzalski's argument against this claim (Gruzalski, pp. 775–76) seems to me to have some force.

APPENDIX E

KUFLIK'S OTHER CLAIMS

Kuflik suggests that, on Common-Sense Morality, we should never give priority to the interests of our own children. Our duties to our children may, he thinks,

be overridden whenever we could do somewhat greater good elsewhere (Kuflik, pp. 787–89). Though this is a purely factual question, I am sure that this is not what most of us believe. Kuflik then suggests that, if M does tell us to give priority to our own children, this is true only in "a very small subset of what is arguably a very small set of cases to begin with" (Kuflik, p. 789). He must be assuming here that, on Common Sense Morality, we should at most give *slight* priority to our own children. But most of us believe that our duties to our children are much greater than our duties to strangers. And we have similar beliefs about many other special obligations. We believe, for example, that a government's duties to its own citizens are much greater than its duties to aliens. My argument therefore applies to many cases (pp. 98, 59–62).

In the cases which Kuflik discusses, each of us can either (E) benefit his own children or (A) give greater benefits to the children of others. I claimed that, in many of these cases, M would tell us all to do E rather than A. Kuflik objects that, for this to be true, M would have to be the "amoral" view that we should benefit our own children *whatever the cost to others* (Kuflik, pp. 785–86, 794, 796, 800, 801). As Kuflik writes, this is not the common-sense view. Most of us would claim that we should not give absolute priority to the interests of our own children, nor should we benefit our children in ways that are unfair.

These claims would not prevent M from telling us to do E rather than A. If we all do E there would be no unfairness. And, in telling us to do E, M need only claim that we should give *some* priority to the interests of our own children. M might claim that we should not do E if the benefit produced by A would be *very much* greater. But this could leave many cases in which, on M, we ought to give priority to our own children.

I claimed that, in these cases, M is directly collectively self-defeating. Kuflik takes my claim to be that, if we all follow M, we will cause our M-given aims to be worse achieved. If this is true, he writes, the problem is not with M but merely with the fact that we cannot communicate, and thereby coordinate our acts. And he claims that this is no objection, since most moral theories, if they are "burdened by the stipulation that people cannot communicate . . . will be unable, in at least some cases, to guide their adherents to (what the theory itself identifies) as the best possible result" (Kuflik, pp. 796–98).

This was not my objection. If M is self-defeating in the way I claim, the problem is not that M fails to *guide* M-believers to the outcome which will best achieve their M-given aims. The problem is that M *forbids* M-believers to produce this outcome. Kuflik overlooks my distinction between a theory's failing to solve coordination problems, and its being directly self-defeating (pp. 53–56). If M is directly self-defeating, the problem *is* with M, not with the fact that we cannot communicate. (This is shown by this problem's disappearance on M's revised version R. Even when M-believers cannot communicate, following R would lead them to the outcome where the M-given aims of each are best achieved.)

APPENDIX F

COORDINATION PROBLEMS

I have claimed that Common-Sense Morality ought to include R because, if it did not, it would be directly self-defeating. This argument can be expressed in a milder way. If M did not include R, Parent's Dilemmas would be coordination problems which M would be, objectionably, failing to solve.

		You: do X	You: do Y
I: do X		Equal-best	Bad
I: do Y		Bad	Equal-best

FIG. 1

Kuflik claims that, if a moral view fails to solve coordination problems, this is no objection to this view (Kuflik, pp. 796–98). But this is true only in some cases.

Suppose that you and I cannot communicate. As we know that we both know, our acts would have the outcomes shown in figure 1. A moral view cannot be expected to solve this problem. Since we cannot communicate, we may have bad luck. I may aim for one of the two best outcomes, while you aim for the other, like two people trying but failing to meet.

Suppose instead that, as we know that we both know, the outcomes would be as in figure 2. A moral view ought to solve *this* problem. Though we cannot communicate, it is obvious that we should both do X. Surprisingly, some Consequentialist theories fail to tell us this. These theories clearly need to be revised.[58]

Parent's Dilemmas are, for Common-Sense Morality, cases of this second kind. As we could know that we all know, if we all do A rather than E, this will be better for all our children. As I argued, this is a better outcome not merely in Consequentialist terms, but in M's terms. If we all do A rather than E, we will thereby cause the M-given aims *of each* to be better achieved (pp. 103–8). Since this is a better outcome in M's terms, M ought to tell us to achieve this outcome. M would tell us this if it includes the whole of R—Rii as well as Ri. This is enough to show that M ought to include the whole of R.

		You: do X	You: do Y
I: do X		Best	Bad
I: do Y		Bad	Second-best

FIG. 2

58. D. Regan, *Utilitarianism and Cooperation* (Oxford: Clarendon Press, 1980).

APPENDIX G

KAGAN'S THEORY

On my account, the Self-interest and Present-aim Theories disagree about the rationality of aims. In telling us to pursue self-interest, S implausibly singles out, and claims to be rationally supreme, one particular aim. This is not true of P. If we accept a pluralistic view about the rationality of aims, we should therefore accept P rather than S.

Kagan rejects these claims. He might say, "S does *not* single out one particular aim. It tells us to try to achieve *all* of our aims, whatever these may be. S and P disagree, not about the rationality of aims, but about the significance of time. While P tells us to try to achieve only our present aims, S tells us to give as much weight to our past and future aims."

This description of the disagreement seems to me mistaken. S tells us to do whatever would be best for ourselves. This is the same, Kagan assumes, as whatever would best fulfill all of the desires that we ever have. But these are the same only on one particular theory about self-interest: the Unrestricted Desire-Fulfillment Theory. As I argued (pp. 494–95), we should reject this theory. We have many desires whose fulfillment is irrelevant to our own self-interest. Suppose that, after a stranger tells me her ambitions, I strongly want her to succeed. If I continue to have this desire, and it is later, without my knowing this, fulfilled, this will not make *my* life go better. It will not affect my experiences; nor will it be, in any other way, good for me.

On the theory which Kagan calls S—or, for short, *Kagan's theory*—I should try to fulfill all of my desires. Since the fulfillment of some of these desires is in no way good for me, this is not a version of the Self-interest Theory. Nor would it make a difference if, as Kagan suggests, his theory told me to ignore those of my desires which are irrational, or would not survive reflection. It is not irrational for me to want the stranger to succeed; and this desire would survive reflection. Why is the fulfillment of this desire neither good nor bad for me? Not because it is irrational, but because it is not a desire about how my own life goes, or whether *my* activities succeed. We have many similar desires.

Kagan suggests that, even if his theory is not a version of S, it is "an important rival to the Present-aim Theory" (Kagan, n. 4). It makes the distinctive claim that, while a rational agent need not be neutral between different people, he must be temporally neutral. But Kagan gives no argument for this claim; nor—I believe—does he answer my arguments against it. Commenting on my analogy between *I* and *now*, he cites irrelevant analogies between *I* and *others*, and between *now* and *then* (Kagan, p. 753). My analogy, though not itself an objection to Kagan's theory, suggests objections. For example, if I can give special weight to *my* aims, why can I not give special weight *now* to my present aims? Kagan admits the force of this objection. On his theory, I should give as much weight to my past and future aims. Kagan admits that, in its claim about my *past* aims, this theory is hard to believe (Kagan, p. 757). It may seem more plausible in its claim about my future aims. But this is because, in many cases, the fulfillment of these aims will be good for me. In these cases Kagan's theory coincides with S and therefore needs no separate discussion. Kagan's theory makes a distinctive claim only when the fulfillment of my future aims will *not* be good for me. Should I give as much weight now to the fulfillment of *these* aims? If I do not have these

aims, and know that their fulfillment will not be good for me, this is not plausible. Another objection to Kagan's theory is provided by those aims which rest on value judgments or ideals. As I wrote, we must give priority to what we *now* value or believe (pp. 153–56). Since I was discussing *subjective* rationality, Kagan's reply (n. 11) fails. I then claimed that most of us are biased towards the future, and would find it hard to believe that this bias is irrational (pp. 158–86). Kagan replies that his theory need not criticize this bias (Kagan, p. 757). For a reason that I gave (p. 132), this reply seems to me inadequate.

Kagan's theory is in one way more defensible than S. It does not insist that we should be governed by any one particular aim. Theory S in contrast does insist that there is one supremely rational aim: our own long-term self-interest. None of Kagan's arguments support this claim.

APPENDIX H

INTERTEMPORAL INJUSTICE

On the argument in my text, since there cannot be compensation over time, we should aim for a fair distribution not between different people but between the different parts of all these people's lives. There are various ways in which this conclusion needs to be qualified.

Suppose that, for the sake of benefits when I am young, I impose burdens on myself in old age. Am I being unfair to my future self? If this seems implausible, this may be because I am not affecting other people. We may think that such an act cannot be unfair because it cannot be morally wrong. Let us therefore consider acts which do affect other people.

How others should be treated should often be for them to decide. But this may be impossible. Thus, if the choice of a social policy would affect many different people, this choice cannot be made by each of these people. It is to such decisions that distributive principles most naturally apply.

There is another complication we should set aside. Resources often produce greater benefits if they are equally distributed both between different people, and between the different parts of these people's lives. Similarly, if we know that our burdens will later benefit ourselves, this may make these burdens easier to bear. It will be clearer to consider cases which do not involve such compensatory effects.

When we receive benefits, they are thought to give us *pure* compensation for our burdens. This is the kind of compensation which I discussed. According to most distributive principles, it has great importance. Thus if we have to bear burdens merely to benefit other people, this would often be thought to be unfair; but if these burdens will be for our own good this would not be thought to be unfair.

If there cannot be compensation over time, as my argument suggests, we should change the scope of distributive principles. According to one such principle, we should give some priority to helping the people who are worse off. On my argument, we should give such priority not to those who are worse off in their lives as a whole, but to those who are worse off at particular times.

This distinction cannot always be drawn, since some things are bad for us without being bad at particular times. One example might be the failure of our ambitions. But other things, such as pain, are bad only at the time. Suppose that

we must choose to which of two people to give some anaesthetic. The first person's pain is now worse, but it is the second person who, in his life as a whole, is worse off. If the principle of equality applies to whole lives, it tells us to help the second person. But it is not absurd to claim that, since there cannot be compensation over time, we should help the person whose pain is now worse. On this view, it is irrelevant that this person has been and will later be better off, since benefits at other times cannot now give him compensation.

In this case, if we help this person, we would also be doing what would more effectively reduce the total sum of suffering. The time-relative version of the principle of equality coincides with the principle of utility. Consider next a case where these principles conflict. Suppose that we must choose to which of two people to give some treatment. If these people are not treated, the first would be in pain for a year, and the second in less severe pain for several years. Though the second person would, in his life as a whole, be worse off, the first person would for a year be worse off than the second person would ever be. If there cannot be compensation over time, the principle of equality tells us to give some priority to helping the first person. Once again, this conclusion is not absurd.

Now suppose that the case involves, not two people, but two alternatives for one person. After the first of two treatments, this person would be in pain for a year; after the second, he would be in lesser pain for several years. Would the first treatment be unfair to him during his worst year? If he prefers this treatment, this view may seem absurd. We might say, "What should happen to this person is entirely for him to decide. Each of us has the right to choose, at any time, how he should be treated at other times."

Is this so? If someone chooses to bear a burden now so as to benefit himself later, we should perhaps treat him now as he chooses. He is like someone who accepts a burden to benefit someone else. But what if the timing is reversed: what if he chooses a benefit now at the cost of a burden later? If there cannot be compensation over time, this is like choosing to benefit himself at the cost of burdening someone else.

In such a case, we should ask whether this person's preferences would later change. If they would not, we should treat him as he prefers. But suppose that, when he comes to bear his burden, he would regret his earlier decision. If he would at different times have such conflicting preferences, we may deny that he can, at any time, make decisions for himself at all future times. Why should his earlier preference have such priority? It may be said that, in making his decision, he commits himself not to complain later if we treat him as he now prefers. But, if he would later regret his decision, this commitment may not be enough to justify our act. Consider a young man who does not now care about himself in old age, and who therefore decides not to make payments to a pension fund, or to medical insurance. When he is old, sick, and poor, he regrets this decision. In such cases we may doubt that someone's earlier decision settles what his fate should be. Treating him as he now wants may be unfair to this person later.[59]

There is another way in which my argument needs to be qualified. I assumed that we cannot be compensated by benefits to other people. If we love these other people, this might be denied. Suppose that we must again choose whom to help. The person who is worst off—call him X—wants us to help, not him, but his children. If X knows that, by bearing a burden, he would benefit his

59. See the chapter on Paternalism in Wachsberg.

children, this might make him glad, which would be good for him. But we should again ignore such compensatory effects. Suppose that X would never know that his burden would benefit his children. It is merely true that, if he did know, he would want to bear this burden.

If X does bear this burden, would the benefits to his children be good *for him?* On one theory about self-interest, the Unrestricted Desire-Fulfillment Theory, the answer is yes. But, as I argued, we should reject this theory. What someone would want, even if he was well-informed and rational, may not be the same as what would be good for him. In deciding what would be good for someone, we should at most appeal to the Success Theory, which gives weight only to this person's desires about his own life (pp. 494–502). On this theory we might claim that, if X's burden benefits his children, these benefits are good for him, since they make him more successful in his attempts to be a good parent. But this claim is controversial. It is better merely to claim that, because X loves his children, the benefits to them give him a kind of compensation. Since it will not be he who receives these benefits, I shall call this "quasi-compensation." We can leave it an open question whether such compensation would be good for him. (Note one difference between this and pure compensation. Future benefits to us will give us pure compensation even if we do not now care about ourselves in the future. But we are quasi-compensated by benefits to others only if we care about these other people.)

What is the moral importance of quasi-compensation? Suppose that, because we are choosing between two social or economic policies, we are forced to decide whether X should bear the burden that would benefit his children. Would it be unfair to impose on him this burden? On one view, we cannot be treating someone unfairly if we are doing what he would want. But when there is no actual consent, the fact that a person *would* consent may not be enough. We may believe that, whatever X would want, we ought to give him his fair share of resources. On this view, though he can redistribute his share, we should not do this on his behalf.

This view seems to be assumed by most egalitarians, since they ignore quasi-compensation. Is this merely an oversight? Should they treat quasi-compensation as they treat pure compensation? This would make the units for distributive principles not different persons but the different groups about whom each person is especially concerned. In many societies, for example, the share of resources received by women is much less than the share received by men. Most of these women care greatly about the well-being of certain men, such as their husbands, fathers, or sons. If the units for distributive principles were the groups about whom each person cares, there would be less ground for claiming that in these societies the distribution of resources is unfair. The lesser shares received by women would be offset by the greater shares received by the men about whom they care. As this example shows, if distributive principles were revised in this way, they would be less plausible. Egalitarians should not give to quasi-compensation the kind of weight that they give to pure compensation.

Return now to distribution within single lives. On the argument in my text, there cannot be pure compensation over time. But there could be quasi-compensation. Just as we can be quasi-compensated by benefits to those other people about whom we care, we can be quasi-compensated now by benefits to ourselves in those other parts of our lives about which we now care. And, even if we became Reductionists, we would still care about our own future. Could such quasi-com-

pensation have the same importance as pure compensation? Such a claim did not seem plausible in the case of quasi-compensation over different lives. The same is true, I believe, within lives. But this is a difficult subject, which needs further thought.

On my argument, if we believe that Backup was not compensated by my luxury, we should conclude that there cannot be pure compensation over time. If we continue to give weight to distributive principles, we should therefore aim for a fair distribution between the different parts of all our lives. I have suggested three ways in which this conclusion should be qualified. Since we care about ourselves at other times, the different parts of each life are like the members of a group who all care about each other. Fair distribution within such groups is much easier to achieve than fair distribution between the mutually indifferent members of a mass society. Similarly, though we may have claims to equal shares at each time, it is not unfair if we unwillingly bear burdens so as to benefit either other people, or ourselves at other times. And, when we apply distributive principles, we should perhaps give some weight to what I have called quasi-compensation. These qualifications make my conclusion less extreme, but easier to accept.