Temporal Neutrality and the Bias toward the Future*

Samuel Scheffler
New York University

1. Introduction: Rawls and Sidgwick on Temporal Neutrality

The conviction that rationality requires an equal concern for all parts of one’s life marks a rare point of agreement among leading Kantian and utilitarian philosophers. John Rawls disagrees with Henry Sidgwick about many things, but the rationality of temporal neutrality is not one of them. In *A Theory of Justice*, Rawls makes his agreement with Sidgwick on this point explicit. He writes:

> In the case of an individual the avoidance of pure time preference is a feature of the rational. As Sidgwick maintains, rationality implies an impartial concern for all parts of our life. The mere difference of location in time, of something’s being earlier or later, is not in itself a rational ground for having more or less regard for it. Of course, a present or near future advantage may be counted more heavily on account of its greater certainty or probability, and we should take into consideration how our situation and capacity for particular enjoyments will change. But none of

---

* This paper was originally written for the *Conference in Memory of Derek Parfit* that was held at Rutgers in December, 2017. I am grateful to the Rutgers audience as well as to audiences at MIT and the University of Pittsburgh for helpful discussion. I am also indebted to Jeff McMahan, Jake Nebel, and Joseph Raz for written comments on earlier drafts.
these things justifies our preferring a lesser present to a greater future
good simply because of its nearer temporal position.

Sidgwick, for his part, thinks that the principle “of impartial concern for all parts of our
conscious life,” as he calls it, represents a commonsense consensus. He says that an
“equal and impartial concern for all parts of one’s conscious life is perhaps the most
prominent element in the common notion of the rational – as opposed to the merely
impulsive – pursuit of pleasure.”

On this last point, at least, Sidgwick was almost certainly mistaken. The
principle of impartial concern may sound commonsensical, but most of us have robust
preferences that are inconsistent with it. That is one lesson of Derek Parfit’s discussion
of temporal bias in Reasons and Persons.

---

1 John Rawls, A Theory of Justice (Cambridge, MA: Harvard University Press, 1971), pp. 293-4. Rawls does not distinguish here between tensed and un-tensed temporal preferences. That is, he does not distinguish between a preference for having certain things occur either in the past or in the future and a preference for having certain things occur either earlier or later. The distinction is significant, however. The bias toward the future is not the same as a bias toward the later, and some un-tensed preferences may be compatible with temporal neutrality, if such neutrality is understood as an impartial concern for all parts of one’s life. See pp. 16-17 below.


3 Ibid., p. 124n.

4 Derek Parfit, Reasons and Persons (Oxford: Clarendon Press, 1984), especially Chapter 8. As Tom Hurka has pointed out to me, Parfit’s position was anticipated by J.M.E. McTaggart, who identified the phenomenon of future bias very clearly in The Nature of Existence, Volume 2 (Cambridge: Cambridge University Press, 1927), Chapter LIX.
2. Parfit on the Bias Toward the Future

In Parfit’s famous example, you are asked to imagine that you awake in your hospital bed. You are told by a nurse that either you underwent ten hours of extremely painful surgery without anesthesia yesterday, after which you were given a drug to cause you to forget the experience, or else you will undergo one hour of extremely painful surgery without anesthesia today, after which you will be given a drug to cause you to forget the experience. Which would you prefer? Most of us would prefer to have undergone the longer operation yesterday than to be faced with the shorter operation today. Using this and other examples, Parfit argues that we display a bias toward the future in the following respects. We would prefer to have experienced pain of a given intensity and duration in the past than to experience it in the future. We would even prefer to have experienced a longer period of pain in the past than to experience a shorter period of pain in the future. And we would prefer that our lives contain more total hours of pain, if that meant less of it were still to come.

With respect to pleasurable sensations, we have the reverse preferences. We would prefer to experience pleasure of a given intensity and duration in the future than to have experienced it in the past. We would even prefer to experience a shorter period of pleasure in the future than to have experienced a longer period in the past. And we would prefer our lives to contain fewer total hours of pleasure, if that meant more of it were still to come. Taken together, and setting aside various refinements and qualifications, these claims ascribe to us a general preference that our pains be in the past and our pleasures in the future. With respect to pleasure and pain, we are, so to speak, more solicitous of our futures than of our pasts. In that sense, we are biased toward the future.
As Parfit makes clear, our attitudes toward time form a complex network, of which the bias toward the future is just one element, and it is not immediately apparent how the various elements fit together. What is evident is that care must be taken in delineating the scope and limits of the bias and its relations and interactions with various of our other temporal attitudes. Three things that Parfit says deserve special attention in this connection.

First, he says that the bias toward the future is not the only form of temporal bias we exhibit. Many of us also exhibit a bias toward the near. We would prefer a smaller pleasure sooner to a larger pleasure later, and a larger pain later to a smaller pain sooner. To the extent that we exhibit both of these temporal biases, there are questions about the relations and possible interactions between them. And in assessing the rationality of the bias toward the future, we must be sure that our assessment is consistent with our assessment of the bias toward the near.

Second, Parfit indicates that the bias toward the future is limited in scope. It applies to certain of our experiences, and especially to our experiences of pleasure and pain, but it does not apply to all the good and bad things in our lives. For example, he says, it is not a feature of our attitudes toward “events that give us either pride or shame; events that either gild or stain our picture of our lives.” Although Parfit may here be underestimating the extent to which experiences of pleasure or pain can themselves be the objects pride or shame – think, for example, of people who feel shame about certain experiences of sexual pleasure or pride about having endured certain

---

1 Indeed, as I have argued in Why Worry about Future Generations? (Oxford: Oxford University Press, 2018, pp. 124-129), there are many good and bad experiences to which it also does not apply.

2 Reasons and Persons, p. 160.
episodes of pain and discomfort – it is not difficult to see the sort of thing he has in mind. Would you prefer already to have written five excellent novels or to write one such novel in the future? Would you prefer to have betrayed your friends five times in the past or to do so once in the future? If our bias toward the future applied to these cases, then we would prefer to write one excellent novel in the future rather than to have written five in the past. And we would prefer to have betrayed our friends five times in the past rather than to do so once in the future. But it is not clear that most people would have these preferences.

Third, Parfit says that the scope of the bias is limited in another way as well. It does not, he asserts, manifest itself as clearly in third-person cases as it does in first-person cases. Imagine you had been informed that your elderly mother, who was living in a distant land and about whom you seldom received news, was suffering from a fatal illness and, moreover, that she was going to have to undergo several months of severe pain before succumbing to the illness. Imagine that you then received a second communication informing you that the initial report had been mistaken about the timing of your mother’s ordeal. In fact, she had already undergone several months of severe pain and had then died. Would you be greatly relieved to learn that her suffering, instead of being in the future as you had initially supposed, was instead in the past? Parfit doesn’t believe that you would. More generally, he writes:

My examples reveal a surprising asymmetry in our concern about our own and other people’s pasts. I would not be distressed at all if I was reminded that I myself once had to endure several months of suffering.
But I would be greatly distressed if I learnt that, before she died, my mother had to endure such an ordeal. This seems to show that the bias toward the future is not as evident in our attitudes toward other people’s experiences as it is in our attitudes toward our own experiences. Although it is a great relief to think that our own pains lie in the past rather than the future, it is not a comparably great relief to think this about those we love.

Despite these complications, the bias toward the future appears to be quite robust. Virtually everyone seems to exhibit the bias. And it does not appear to be subject to direct volitional control. That is, it does not appear to be something one could simply decide to give up. Caspar Hare says that “you may as well try to lose your skin.” But while this casts doubt on the commonsense credentials of Sidgwick’s principle “of impartial concern for all parts of our conscious life,” it doesn’t show that the principle is mistaken. It doesn’t show that temporal neutrality is not a requirement of rationality. It may be that the bias toward the future, widespread and robust though it is, is nevertheless irrational. Perhaps we are all subject to the bias, but perhaps we are all, to that extent, irrational.

What does Parfit think? Does he mean to be claiming not only that the bias is deep and widespread but also that it is rationally required or at least rationally

\*\*\*\*

\* Ibid., p. 182.

\* Caspar Hare, “Time – The Emotional Asymmetry,” in Heather Dyke and Adrian Bardon eds., A Companion to the Philosophy of Time (John Wiley & Sons, Inc., 2013): pp. 507-20, at p. 513. Compare McTaggart: “[While] allowing that past good or evil does tend to produce present happiness or unhappiness, it remains the case that future good or evil tends to produce them to a much greater extent. I do not know that any reason can be assigned for this greater present importance of the future. It may have a reason which is not yet discovered. Or it may be an ultimate fact. But it cannot be denied to be a fact” (The Nature of Existence, Volume 2, p. 350).
defensible? Or does he, in the end, side with Sidgwick and Rawls, and suppose that the bias is irrational and that temporal neutrality is, after all, a requirement of rationality? In fact, Parfit’s position is surprisingly difficult to pin down. This is largely because of the dialectical context in which his discussion of the bias toward the future is embedded. He raises the issue in the course of an extended argument against what he calls S, or the Self-interest Theory. And rather than assessing the bias directly, or in his own voice, often what he does instead is to draw conclusions about what the S-theorist must say about the bias. For example, he writes:

If [the S-Theorist] condemns the bias toward the near because it cannot have rational significance when some pain is felt, he must condemn the bias toward the future. He must claim that it is irrational to be relieved when some pain is in the past. Most of us would find this hard to believe. If the S-Theorist insists that we should be temporally neutral, most of us will disagree.”

In this passage, Parfit seems to be treating it as an implausible implication of S that, on certain assumptions, it must condemn the bias toward the future as irrational. So it is natural to conclude that he himself thinks the bias is not irrational. And that is how some readers have interpreted him.” But he never actually says this.

---

¹ In fairness to Rawls, it should be noted that, when he says that temporal neutrality is a requirement of rationality, it appears that his main concern is to insist on the irrationality of the bias toward the near. He does not explicitly mention the bias toward the future, and it is not clear whether he is taking account of the fact that the requirement of temporal neutrality condemns both biases.

² Reasons and Persons, p. 170.

³ Some others have interpreted him as being neutral about the rationality of future-bias. See, for example, Preston Greene and Meghan Sullivan, “Against Time Bias,” Ethics 125 (2015): 947-970, at 967n.
Nor does he say explicitly in *Reasons and Persons* that the bias toward the future is irrational. But he does say, somewhat surprisingly in light of passages like the one just cited, that S’s requirement of temporal neutrality is “plausible,” and that “this part of S may be either true or, if we think that such claims cannot be true, part of the best or best justified theory.” And he devotes a section to arguing that we should not be biased toward the future, because the bias is bad for us. In developing this argument, he asks us to consider a person he calls *Timeless*, who lacks the future-bias. For Timeless, looking backward to past enjoyments is just as pleasant as looking forward to future ones, and looking backward to past pains is just as distressing as looking forward to future ones. If we were like Timeless, Parfit says, things would be in some ways worse for us. For example, we would not be relieved when bad things were in the past. On the other hand, we would not be sad when good things were over. In addition, we could afford to be selective about which events we chose to look backward to. We “could allow ourselves to forget most of the bad things that have happened, while preserving by rehearsing all our memories of the good things.” So the net effect of these changes would be positive. Furthermore, we would gain enormously in our attitudes toward our deaths. Although, as death approached, we would have less and less to look forward to, we would have more and more to look backward to. On our

— *Reasons and Persons*, p. 194.

Parfit seems to understand *looking backward* to past enjoyments as not requiring memory of those enjoyments. Although, as he emphasizes, the memory of a pleasurable experience can itself be pleasurable, and although Timeless experiences this sort of pleasure, he also enjoys knowing that he had pleasant experiences even if he has no memory of them. Or at any rate, that is how I interpret the relevance of Parfit’s discussion in the final complete paragraph on p. 172 of *Reasons and Persons* for his subsequent characterization of Timeless on p. 174.

— *Reasons and Persons*, p. 175.
deathbeds, we would have our whole lives behind us, and that would be almost as good as having our whole lives ahead of us."

3. Is the Bias Toward the Future Bad for Us?

However, this accounting of the advantages and disadvantages of Timeless’s situation – or of what I will call *Timelessness* – is not comprehensive. There are several points to consider. First, Parfit says that Timeless is never relieved when bad things are in the past. When he “is reminded that he once had a month of agony, he is as much distressed as when he learns that he will later have a month.” Now to me this sounds like a prescription for universal PTSD. Parfit’s belief that it would merely be a disadvantage that would be outweighed by the various advantages of Timelessness rests on the thought that we could choose to preserve our memories of the good things that have happened and to “allow ourselves to forget” the bad things. But this assumes without argument a strongly voluntaristic conception of memory. Moreover, even if we make generous assumptions about people’s ability to choose which past events to dwell

"Why “almost”? Parfit’s reasoning may seem to show that it would be *just* as good. But there is a complication: “This reasoning ignores those emotions which are essentially future-directed. It would not apply to those people for whom the joy in looking forward comes from making plans, or savouring alternatives. But the reasoning seems to be correct when applied to more passive types, those who take life’s pleasures as they come.” Accordingly, Parfit’s more guarded conclusion is that, to the extent that we are like these more passive types, “we would be happier if we lacked the bias towards the future. We would be much less depressed by ageing and the approach of death. If we were like Timeless, being at the end of our lives would be more like being at the beginning. At any point in our lives we could enjoy looking backward or forward to our whole lives” (*Reasons and Persons*, 176-7).

on when engaged in private contemplation, and even if we ignore questions about the status of unconscious memories and the possible psychic costs of suppressing them, there is also the role of other people to consider. For example, other people’s sufferings often trigger memories of one’s own, and Timelessness would ensure that such memories were just as distressing as an awareness of one’s future sufferings. Furthermore, the fact that reminding someone of his past suffering could produce extreme distress suggests new possibilities for malicious behavior. In order to torment someone one disliked, one would need only to issue such reminders. (“Remember the time you got food poisoning after eating bad seafood, or the day your partner left you, or the time you made a fool of yourself at the party, or the day you were denied tenure?”) For Timeless and those like him, time heals no wounds, and the only alternative cure would be a form of selective amnesia so powerful that even the interventions of other people could not revive the forgotten memories.

Second, it is worth wondering what the effects of Timelessness would be on human motivation. Once one had lived long enough to have accumulated some good experiences, would one still have the same motivation to seek out additional good experiences, with all the attendant effort, uncertainty, and risk of failure that involves, when one could just look back with pleasure to the old ones instead? Of course, Timelessness would not mean that looking back was bound to be just as pleasant as the new experiences themselves might be, only that it would be as pleasant as looking forward to those experiences. But, of course, the new experiences themselves would have no advantage as such over the old experiences, and so the question is how, if looking forward also had no advantage over looking backward, this would affect the structure of our motivations and our incentives to seek out new experiences. To put it another way, what are the cognitive, motivational, and developmental roles played by
attitudes like *anticipation* in our psychic economy, how would Timelessness affect those roles, and with what consequences? (Parfit himself seems undecided, in his discussion on pp. 172-4 of *Reasons and Persons*, between a thin conception of anticipation as consisting simply in the thought that something will happen and some thicker, though not fully specified, conception.) I don’t suppose that these questions can be answered from the philosophical armchair. But they do need to be answered if we are fully to assess the advantages and disadvantages of Timelessness for creatures like us.

Third, there is also the question of how human relationships would be affected by the elimination of the bias toward the future. I have already suggested that Timelessness would open up new possibilities for malicious behavior. Equally, of course, it would also open new possibilities for kindness: one could reliably cheer people up by reminding them of good times they’d experienced in the past. But malice and kindness aside, there is a more general question about the effects of temporal neutrality on personal relationships. Consider the simple fact that we often look forward to seeing our friends and family members in the future. Timelessness would mean that it would be just as satisfying to look backward to having seen them in the past. How would this affect people’s interactions and their desires to spend time together? How would it affect the structure of human attachments and our responses to loss? Again, I don’t pretend to know the answers to these questions, but I am reluctant to draw conclusions about the net advantages or disadvantages of Timelessness without knowing them.

Fourth, consider the appetites. Parfit’s discussion may create the impression that Timeless looks backward to, say, past gustatory pleasures with just as much enjoyment as he looks forward to future gustatory pleasures. But this is misleading. When I am hungry, it is not just my future-bias that leads me to look forward to eating my next
meal more avidly than I look backward to eating my last one. There is also the fact that my next meal can satisfy my present appetite, while my last meal cannot. I assume that Timeless is no different from me in this respect. When he is hungry, he too looks forward to eating his next meal with greater enthusiasm than he looks backward to eating his last one. But if that is right, it suggests that Timeless’s neutrality with respect to past and future appetitive pleasures is narrower in scope, and its advantages are therefore more limited, than we might initially have assumed.

Fifth, Parfit says that, as death approached, the fact that Timeless would have less and less to look forward to would be offset by the fact that he would have more and more to look backward to. But of course, this works only if Timeless has accumulated a sufficient store of good experiences to look back to. Those people whose lives have been filled with suffering would have little to look forward to and little to look backward to, so for them the advertised advantages of Timelessness would be elusive. The point is not that these people would be better off if they retained a bias toward the future. The point is rather that the alleged advantages of Timelessness, with respect to our attitudes toward aging and death, would not be available to everyone. They would accrue only to those who had been fortunate enough to lead reasonably good lives.

Sixth, when assessing the effects of Timelessness on our attitudes toward death, it is worth bearing in mind Parfit’s caution that the bias toward the future applies to certain of our experiences, and especially to our experiences of pleasure and pain, but that it does not apply to “events that give us either pride or shame; events that either gild or stain our picture of our lives.” With respect to these events, we exhibit no bias toward the future even now. This means that, as death approaches, we are already in as good a position as Timeless is to take solace from the things we have done that give us pride. We are also, of course, in just as good a position, if that’s the right expression, to
be tormented by the things we have done that give us shame. But many people do derive solace, in the face of their impending deaths, from their justifiable pride in things they have accomplished during their lifetimes. Yet I think it is fair to say that, even for such people, this does not always extinguish the fear of death. Far from it. And so it is important to emphasize that Timelessness would confer no additional advantage with respect to attitudes like pride or shame, or with respect to the things that gild or stain our picture of our lives. The extra advantage it would confer on us, in looking backward, is merely that we might derive greater pleasure from contemplating our past experiences of pleasure. But if even justified pride in one’s past achievements is not enough to banish the fear of death, it is perhaps not obvious that the pleasure of thinking about past pleasures would be more effective.

Finally, Parfit may be underestimating the advantages of our existing attitudes toward death. Consider that people who are aware of the imminence of their deaths can no longer sensibly occupy themselves with trying to shape their lives. To the extent that they are inclined to engage in self-reflection, their gaze must be directed primarily backward rather than forward. Given their biased tendency to discount past pleasures and pains, Parfit thinks they are unlikely to find this enforced backward gaze rewarding. Accordingly, he imagines that it would be a great improvement for them if, like Timeless, they could relish their past experiences just as avidly as they had once relished the prospect of future experiences. But this neglects the fact that the pressure to attend to the past presents a temporally biased person who is nearing death with a distinctive challenge and, perhaps, a distinctive opportunity. The challenge is to come to terms with one’s life as it has actually turned out. As a practical matter, one is no longer in a position to change one’s life in fundamental ways. Yet as an epistemic matter, one is in a better position than ever before to survey the entire trajectory of one’s
life; over time, ignorance of one’s future has been almost entirely supplanted by knowledge of one’s past. Finding oneself in these circumstances – with one’s practical position diminished but one’s epistemic position enhanced – and lacking Timeless’s inclination to dwell lovingly on the hedonic high points of yesteryear, the challenge is simply to reckon with one’s life, in all its awe-inspiring contingency, finitude, and immutability. If one is fortunate, this challenge may bring with it a certain opportunity: the opportunity to achieve a kind of self-acceptance, or to make peace with what one has become. If, by contrast, we were like Timeless, we might just go to our graves looking back with delight on experiences of great meals, great music, or great sex. And I am sure some will agree with Parfit that that would be a vast improvement over the current arrangement. But others may wonder whether the opportunity to take the measure of one’s life, to view it as a (nearly) complete object and to try to make sense of it as such, does not have a value of its own. They may wonder whether the advantages of Timelessness, as death approached, would be quite as clear-cut as Parfit suggests.

Of course, the considerations I have been rehearsing do not show that Parfit is wrong in thinking that the bias toward the future is bad for us and that we would be better off if we were like Timeless. But they do suggest that, in order to arrive at defensible conclusions about the overall advantages and disadvantages of the bias, a more comprehensive accounting would be required.

4. Is the Bias Toward the Future Irrational?

Suppose, however, that Parfit is correct, and that the bias toward the future is indeed bad for us. How exactly does this bear on questions about the rationality of the bias? As I’ve noted, Parfit concludes his discussion by saying that the requirement of
temporal neutrality is plausible and may indeed be true. And he said subsequently that he did mean to be endorsing the view that rationality requires temporal neutrality and that the bias toward the future is irrational. But if that is his view, then what are the arguments in its favor that he finds persuasive? One obvious possibility is that he thinks the bias is irrational because it is bad for us. The difficulty with this suggestion is that Parfit explicitly rejects it. After concluding that the bias is bad for us, he adds that this “does not beg the question about the rationality of this bias. On any plausible moral view, it would be better if we were all happier. This is the sense in which, if we could, we ought not to be biased toward the future.” In other words, we would be happier if we lacked the bias, and so it would be better from a moral point of view if we did, but this leaves open the question of whether the bias is rational. To emphasize the point, he says a few pages later that “the rationality of an attitude does not depend on whether it is bad for us.” But if it is not the (presumed) fact that the bias is bad for us that leads Parfit to think it is irrational, then what are his reasons for that conclusion?

The text of Reasons and Persons leaves this obscure. As I’ve said, many of his arguments are ad hominem arguments directed against the S-theorist. They are about what the S-theorist must say about the bias rather than about the bias itself. And by emphasizing how difficult most of would find it to believe that the bias is irrational, so that it would be awkward for the S-theorist to be committed to that view, those

---


“Ibid., p. 177.

“Ibid., p. 185.”
arguments create the impression – mistaken, as it happens – that Parfit himself thinks
the bias is not irrational, rather than that it is.

Perhaps, like Sidgwick and Rawls, Parfit is simply struck by the force of the
thought that whether an experience is in the future or in the past is not itself a
reasonable ground for having “more or less regard for it.” But nothing has done more
to call the force of that thought into question than Parfit’s own arguments about the
strength and robustness of our bias toward the future. Can we really suppose that
rationality requires us, in Parfit’s hospital case, to hope that we will undergo the painful
surgery today? Given the strong preference that many of us will have to the contrary, a
preference we are likely to affirm upon reflection, more may be required to persuade us
of this than the bare assertion of the rational irrelevance of temporal location.

One argument against the rationality of future-bias appeals to the limitations on
the scope of the bias that Parfit mentions. If the bias does not apply to all goods, and if
it does not apply in the same way to our third-person judgments about other people’s
pleasures and pains, then this may be said to show that it is rationally anomalous.
David Brink, who makes an argument of this kind, observes that our biased preferences
are also diachronically unstable. If I view the options in Parfit’s hospital case from a
temporal perspective where both options lie in the future – say, before I enter the
hospital – then I will prefer the shorter operation to the longer one, even if I know that
shorter operation would take place at a later time than the longer one. So too if I am
asked about my preference when both options lie in the past – as, for example, when I
am leaving the hospital. The bias emerges only when, as in Parfit’s original version of
the example, the options “straddle” the present: when one of them lies in the past and
one of them lies in the future. As Brink says, “it’s not about preferring earlier pain to
later pain; instead, it’s about preferring past pain to future pain.”* This, he says, makes the bias “more narrow or isolated.”* Summing up, he concludes that the bias “does not generalize well and remains limited in scope and unstable.”*

One thing this form of diachronic instability suggests is that, when we are contemplating alternative actions we might perform or experiences we might undergo or things that might happen to us, we can think of the alternatives in two different ways and can ask two different questions about them. Abstracting from our temporal position, we can ask: which alternative would make my life as a whole go better? But viewing the alternatives from a temporally situated perspective, in which it is salient that some of our life lies in the past and some of it in the future, we can also ask: which one would make the rest of my life go better? We can, for convenience, label these two

*David Brink, “Prospects for Temporal Neutrality,” in C. Callender ed., The Oxford Handbook of Philosophy of Time (Oxford: Oxford University Press, 2011), pp. 352-81, at p. 377. This point is significant because some people think that the temporal order of one’s experiences can make a difference to the quality of one’s life. For example, some people think that a life filled with misery in the early years but joy in the later years is better than a life filled with joy in the early years but misery in the later years, even if the amounts of misery and joy in the two lives are the same. They think, in other words, that a life with an upward trajectory is better than a life with a downward trajectory. Similarly, some people think that it is better if satisfactions are distributed evenly throughout one’s life than if they are clustered in one temporal period. These are sometimes described as views according to which the “narrative structure” of a life contributes to its value. (See, for example, David Velleman, “Well-Being and Time,” Pacific Philosophical Quarterly 72[1991]: 48-77, reprinted in J.M. Fischer ed., The Metaphysics of Death [Stanford: Stanford University Press, 1993], pp. 329-357.) The “narrative preferences” just mentioned are not equivalent to, and are not always compatible with, the bias toward the future. And although they may seem to violate temporal neutrality, they do not. Although they assign significance to the temporal location of good and bad experiences, they do so only insofar as the temporal location of those experiences affects the overall value of one’s life from a neutral perspective. For that reason, they are compatible with an impartial concern for all parts of one’s life. See Brink, pp. 357-358.

* Ibid.

different ways of thinking the whole-life perspective and the future-facing perspective, although we should take care not to reify these “perspectives” or to think of them as constituting elements or modules of the self. They are simply two different ways of thinking that are available to us. Most of the time, there is no conflict between them. We will prefer the same alternative no matter which question we ask. The same alternative that would make my future better would also make my life as a whole better. But sometimes the answers may diverge. The bias toward the future manifests itself in cases where such divergence arises and I give priority to the future-facing perspective. When this happens, then we will see the form of diachronic instability that Brink identifies. If I am asked which of Parfit’s two operations I would prefer before I enter the hospital, my preference will be for the shorter operation. From both the whole-life and future-facing perspectives, that is the better alternative. But if, as in Parfit’s original example, I am asked the same question when I know that the shorter operation is in the future and the longer operation is in the past, then my preference will be for the longer operation. From the future-facing but not the whole-life perspective, that is the preferable alternative. Here, in other words, the two perspectives diverge, and I give priority to the future-facing perspective. I display a bias toward the future.

Even if I know all this in advance, the same pattern of preference is likely to persist. That is, if I am asked for my preference before I enter the hospital, I will continue to prefer the shorter operation even if I know that, should I end up in the situation Parfit describes, I will then prefer to have had the longer operation. And if I am asked for my preference in that situation, then I will continue to prefer to have had the longer operation, even if I know that, had I been asked for my preference before I
entered the hospital, I would have preferred the shorter operation. Mere self-awareness does not produce any change in my pattern of preference.

The idea that we can see our alternatives from two different perspectives, and that giving priority to the future-facing perspective in cases where the two perspectives conflict produces diachronic instability, has been developed and extended by other writers, including Tom Dougherty\textsuperscript{23} and Caspar Hare\textsuperscript{24}. Dougherty says that one perspective we take toward ourselves and our experiences is temporal and perspectival, while the other involves seeing ourselves as temporally extended agents. Hare speaks instead of a contrast between our attitudes toward predicaments and our attitudes toward lives. These distinctions are similar to my distinction between the future-facing and whole-life perspectives, and for present purposes we can treat them as different ways of characterizing the same distinction. Both Dougherty and Hare describe complex cases in which the way we respond to divergence between the two perspectives seems to result in cycling or intransitivity of a kind that most people regard as irrational. The problem arises because, when the perspectives diverge, it appears that we don’t always give priority to the future-facing perspective. Sometimes we see things from the whole-life perspective instead. In consequence, we display the bias toward the future in response to some but not all cases of divergence, and it is this that leads to intransitivity. Both Dougherty and Hare believe that the solution, in the cases they describe, is to give one of the perspectives consistent priority over the other. Dougherty believes that, both in his cases and more generally, we should always give


priority to the whole-life perspective. He agrees with those who think that the bias toward the future is irrational. Hare, by contrast, believes that we should always give priority to the future-facing perspective, at least in cases like those he describes.

5. The Tolerant Stance

I will return to the question of intransitivity. But first let me put some of my own cards on the table. I take it for granted that the temporal dimension of human life is, for anyone trying to lead such a life, one of its most mysterious and perplexing aspects. Philosophy aside, we experience ourselves as moving from the past into the future, and we struggle to understand what it means, and what it means to us, to have a past and a future: to be self-conscious creatures extended in time but living always in the present moment. We struggle to understand what our pasts mean to us, what kind of reality they have for us, and what resources, beyond the fragile tool of memory, sustain our ongoing connections to our past. At the same time, we are moving always into a future that presents itself to us as open, in ways that seem both fundamental and elusive, and we are preoccupied with the implications of that openness. We worry and we wonder about what the future will bring, and about how we can and cannot influence it. And we have a whole range of attitudes that we deploy in orienting ourselves to that future: from hope and fear to anticipation and dread to determination and resignation.

Moreover, as we look backward to the past and forward to the future, our attitudes and actions are guided by values – including but not limited to conceptions of what would we be good for us – that are themselves shaped by our self-understanding as temporally extended creatures and by our experience of temporality. We would not have the values we have if we did not understand the temporal dimension of our lives
in the ways that we do. At the same time, our values serve in turn to shape our attitudes toward time. We would not have the temporal attitudes that we have if we did not have the values that we do. In my view, and here I am doing no more than articulating an unargued methodological predisposition, the first task of philosophical reflection in this area is to try to understand this complex terrain, and we should not be too quick to assume that every manifestation of temporal bias in our desires or valuing attitudes is irrational. Indeed, since the very term ‘bias’ imports a certain bias, the tendency of philosophers to use the term in this context is to that extent unfortunate (although, in deference to the prevailing usage, I will continue to do so myself). As with studies of rational judgment and decision-making in other areas, the aim should be to navigate between the complacent assumption that our ordinary thinking must be in good order and the revisionist application of oversimplified models that lack any authority over our actual practices and tendencies of thought.25

Against this background, it may seem tempting to say something like the following about our bias toward the future. It is a basic feature of our practical thought that we can assess our alternatives from both the whole-life and future-facing perspectives. So if it is true that, by virtue of our responses to cases in which the two perspectives diverge, we exhibit certain forms of diachronic instability or are apt to form intransitive preferences in certain contexts, perhaps it is a mistake to try to eliminate this complexity or to insist that one perspective must rationally supersede the other in all cases of conflict. Perhaps we should be less preoccupied with subjecting our responses to rational criticism. Perhaps we should simply accept the fact that, in this respect as in many others, we are complex creatures, and focus our efforts on trying to

---

25 This paragraph is drawn, with some modifications, from my discussion in Why Worry about Future Generations?, p. 121.
understand the roles of the two perspectives in our lives. I will call this the tolerant stance toward future-bias. The tolerant stance urges us to accept both our future bias and its limits: both the fact that we exhibit the bias in some contexts and the fact that we do not exhibit it in others.

A word of clarification is in order here. As I have said, the bias toward the future manifests itself in cases where the whole-life and future-facing perspectives diverge and we give priority to the future-facing perspective. When we do not exhibit future-bias, that may be either a) because the two perspectives do not diverge or b) because they do diverge and we give priority to the whole-life perspective. In particular cases, it may be open to debate which of these conditions applies. For example, consider the case of goods that “gild or stain our picture of our lives.” Suppose I would prefer to have written five excellent novels in the past rather than to write one such novel in the future, and so I do not manifest a bias toward the future in this instance. It might seem that the reason I do not manifest such a bias is that I am giving the whole-life perspective priority over the future-facing perspective. But another possibility is that, in this case, the two perspectives do not actually diverge, because, say, having written an excellent novel is an achievement that benefits a person throughout the person’s life and not merely at the time when it is accomplished. According to this view, I will actually have a better future, as well as a better life, if I have written five novels in the past. The question of which diagnosis is correct, in this case and others like it, raises deep issues which lie beyond the scope of his paper. The tolerant stance is neutral on this question: it urges acceptance of the limits of our future bias however those limits are explained.²

² I am grateful to Jake Nebel for illuminating discussion of these points.
6. The First Obstacle to Endorsing the Tolerant Stance: Near-Bias

Even for someone of my methodological predisposition, however, there are at least two obstacles to endorsing the tolerant stance. To appreciate the first obstacle, consider the other temporal bias that Parfit discusses, the bias toward the near. Manifestations of the bias toward the near also lead us to exhibit various forms of instability in our preferences. Yet it seems clear that manifestations of this bias can indeed expose us to rational criticism. But if that is right, then more needs to be said to explain why the bias toward the future is different, and why it should be tolerated or accepted rather than criticized.

One difference lies in our own attitudes toward the two biases. As I have noted, most people are strongly inclined, even upon reflection, to reaffirm their future-biased reactions in cases like Parfit’s hospital example. And few of them feel that they are making a mistake in doing so. Instead, their preference to have had the longer operation yesterday rather than the shorter operation today continues, with one qualification that I will mention later, to make sense to them. Even upon reflection, in other words, most people believe that the preference to have had the longer operation yesterday makes sense in the situation Parfit describes. This is, of course, compatible with recognizing that that preference is diachronically unstable. To reflectively reaffirm the preference is not to deny that one would have different preferences if both options

were either in the past or in the future. Those preferences too seem appropriate upon reflection.

With the bias toward the near, things are different. Manifestations of this bias are often subject to criticism by the very people who manifest it. Even when our bias toward the near leads us to postpone dentist appointments, fail to save for retirement, and avoid doing chores today that will only be more difficult to do tomorrow, often we feel uneasy about what we are doing and regard it as revealing a weakness or flaw or vice of some kind. Moreover, many people try to resist this bias, and many people are successful. They do undergo routine dental care and save for retirement, and they rarely put off until tomorrow what they can do today.

What might explain this difference in our attitudes toward the two biases? Consider the case of pain, and let me begin with some undigested phenomenology. Phenomenologically speaking, both biases reflect a concern to keep pain at bay. We think of our present consciousness as moving forward into the future and, as we move forward, we are concerned to avoid having any pain impinge on that consciousness. We might call this our defensive impulse. Since we experience our conscious selves as moving forward in time, the defensive impulse is satisfied by the recognition that an episode of pain is in the past. Past pains will never again impinge directly on our consciousness (which is not to deny that the memory of pain can itself be disturbing). They are over and done with. Future pains, however, pose a threat, since we will eventually cross paths with them; they will indeed impinge on our consciousness. So,

---

It is noteworthy that, while many people have proposed psychological or evolutionary explanations for each of the two biases considered separately (see, for discussion, Greene and Sullivan, “Against Time Bias,” pp. 966-970), the availability of such explanations by itself does not explain the difference in our own attitudes toward the two biases.
from a defensive standpoint, it makes perfect sense to prefer past pains to future pains. Future pains pose a threat in a way that past pains do not. That is why future-bias makes sense to us. The bias toward the near, meanwhile, also has defensive roots. By postponing our pains, we keep as much temporal distance as possible between us and them, and this has an obvious defensive point. Hence the bias toward the near. Yet, at the same time, this bias is open to criticism from the defensive standpoint itself. By postponing our pains, we do increase the temporal distance between us and them, but it remains the case that we will eventually encounter them, and, by hypothesis, postponement ensures that when we encounter them, they will be worse than they would otherwise have been. So there is a clear sense in which the bias toward the near, as a defensive strategy, is self-defeating. That is why, even when we are in the grip of the defensive impulse, we may be uneasy about the bias toward the near and may withhold our reflective endorsement of it. We may think of it as constituting a flaw or defect of character.

Just to be clear, the phenomenology I have described presupposes that we experience ourselves as moving forward in time, and that we therefore have asymmetrical attitudes toward past and future. It makes no attempt to vindicate that view. The thought is simply that, given that we experience ourselves that way, we can see why the bias toward the near strikes us as problematic in a way that the bias toward the future does not.

But we need not appeal to phenomenology to account for our different attitudes toward the two biases. We may simply observe that the bias toward the near, unlike the bias toward the future, is not a response to divergence between the whole-life and future-facing perspectives. In cases where we manifest the bias toward the near, we display a preference for an alternative that is condemned by both perspectives. That is,
the alternative we prefer is one that will both make our future go worse and make our life as a whole go worse. No wonder, then, that we are likely to withhold our reflective endorsement of the bias.

The considerations I have been discussing suggest that the bias toward the future is a deeper or more securely-entrenched feature of our attitudes than is the bias toward the near. And they help us to understand how and why that is so. In so doing, they make it easier for someone with my methodological predisposition to endorse the tolerant stance toward future-bias. They address the first of the two obstacles I mentioned to such endorsement.

7. The Tolerant Stance Revisited

Before describing the second obstacle, let me say a bit more about what the tolerant stance amounts to. In suggesting that we should accept our bias toward the future, I am not suggesting that the bias is rationally required or mandatory. There are several reasons why I would not want to say that. First, and most straightforwardly, I simply cannot think of any plausible norm of rationality that might require future-bias. Second, although I don’t think I have ever met anyone who lacked such a bias, I am in no position, from my armchair, to exclude the possibility that there are such people. Perhaps complete temporal neutrality is the sort of thing one can achieve through a

29 The view that the bias is rationally mandatory is strongly suggested by Heathwood, who writes that he is “not convinced that reason permits Parfit to prefer to be the patient whose operation is later today” (“Fitting Attitudes and Welfare,” p. 61). He also writes, “Parfit is being completely reasonable in preferring that his pain be in the past. In fact, even his no longer caring at all that it occurred is perfectly fitting – not at all inappropriate. Why should he care about it now? No reason – it’s over and done with” (pp. 56-57).
sufficiently disciplined and rigorous regimen of training and meditation. Or perhaps there are rare individuals who are just naturally like Parfit’s Timeless. I doubt it, but I can’t rule it out. And if there are people who genuinely lack a bias toward the future, either as a result of undergoing some training regimen or because that is simply their nature, then I am not inclined to say that they are being irrational or to subject them to rational criticism: not, at least, without knowing a lot more about them. My firmest conviction about such people is that, if indeed they exist, they must be very different from the rest of us in many ways, and I would be immensely curious to know what their lives are actually like.

There is a third reason I would not want to say that the bias toward the future is rationally required or mandatory, and it has to do with the qualification I alluded to earlier when asserting that most of us reflectively affirm our future-biased preferences. What I had in mind was that, despite our reflective endorsement, these preferences may nevertheless be sources of puzzlement or bemusement. One needn’t be a professional philosopher to be struck at times by the thought that there is something strange about our attitudes toward our past and future experiential states and about the broader attitudes toward time that they reflect. It can seem mysterious, at least if one is in a certain sort of mood, that we invest future experiential states with such great emotional and attitudinal significance – that we make them the objects of attitudes like fear, dread, longing, and anticipatory excitement – only to dial back the emotional temperature nearly to zero, and to turn most of our attention away from them, once we have actually undergone them. We may wonder: How can they have merited such intense reactions in advance if we find them so forgettable once we have actually experienced them? Why doesn’t our prospective awareness that this abrupt change will happen do anything to diminish the intensity of our anticipatory attitudes? Why doesn’t my
awareness of the fact that I will be indifferent to tomorrow’s pleasure or pain once the following day dawns make me more nearly indifferent today? What does it say about us that we are more focused on future experiences that we have not undergone than on experiences that already belong to our personal histories? What does this say about our relation to our own pasts, about the significance for us of those pasts, and about the sense in which they are ours?

I rehearse these questions to make what is only a limited point. Most of us have future-biased preferences, and we are not inclined to abandon them upon reflection, but they can also strike us as mysterious, and as exemplifying broader attitudes toward time and toward our lives as temporally extended creatures, that are equally mysterious. And this is another reason why, although I am inclined to accept our future-bias, I would be reluctant to say that it is rationally mandatory or required, and the tolerant stance does not in fact say that.

But if the tolerant stance does not view the bias is rationally required, then neither does it view it as a form of human *irrationality* to which we must simply resign ourselves. The acceptance that is part of the tolerant stance is not a world-weary recognition of the inevitable flaws and foibles of human beings. That attitude, or something like it, may be an appropriate response to some familiar forms of human irrationality, such as akrasia, or the bias toward the near, or some of the deliberative anomalies documented by psychologists like Kahneman and Tversky. Many of us do our best on a case-by-case basis to avoid falling prey to those forms of irrationality, but we don’t suppose that human beings will ever cease altogether to exhibit them. So while we resist them on the micro-level, we despair of eliminating them on the macro-level, and in that sense we accept them. But that is not the kind of acceptance that is
urged by the tolerant stance toward future-bias, because the tolerant stance does not condemn that bias as irrational in the first place.

If accepting the bias toward the future amounts neither to thinking that it is rationally required nor to thinking that it is an ineliminable form of human irrationality, then what’s left? I see two possibilities. The most obvious suggestion is that accepting future-bias means judging it to be rationally permissible though not required. The alternative is to say, not that future-bias is rationally permissible though not required, but rather that it is one of those things – like, perhaps, our need for love or companionship, or our vulnerability to boredom – that help to define the type of rational creature we are rather than something that is itself up for rational assessment. I find this suggestion tempting, but to make good on it, one would need to explain what the difference between being rationally permissible and being ineligible for rational assessment really amounts to. Since the issues here are subtle and space is limited, I won’t pursue this question any further. Instead, I will assume that, in urging that we accept the bias toward the future as part of our nature as complex creatures, the tolerant stance is recommending that we think of the bias in one of the two ways just mentioned: either as rationally permissible but not required or as not subject to rational assessment at all.

8. The Second Obstacle to Endorsing the Tolerant Stance: Intransitivity

Let me turn now to the second of the two obstacles I mentioned to endorsement of the tolerant stance. As I said earlier, both Dougherty and Hare describe cases in which, it seems, we respond to divergence between the two perspectives by giving priority to the future-facing perspective at some moments but to the whole-life
perspective at others, and this results in cycling or intransitivity of a kind that is usually seen as irrational. If that is correct, then the tolerant stance, which neither condemns our future-bias nor insists that we must display it in every case of divergence, may seem simply mistaken. If the tolerant stance is to be defensible, something must be said to take the sting out of the cycling arguments.

It is important to remember that what the cases described by Hare and Dougherty appear to show is not that future-bias by itself leads to intransitive preferences, but rather that the disposition to manifest future-bias in response to some but not all instances of divergence between the two perspectives can lead to intransitivity. Dougherty concludes that, in order to avoid such intransitivity, we should give priority to the whole-life perspective in all cases of divergence and should reject future-bias as irrational. Hare, on the other hand, argues that we should display more future-bias, at least in cases like the ones he describes, than most of us already do. The tolerant stance suggests still another possibility, namely, that we need not revise our responses at all. If it is true, as Hare and Dougherty suggest, that we display future bias in response to some but not all instances of divergence between the two perspectives, then we can accept that pattern of response even if it leads to the formation of intransitive preferences in cases like those they describe.

Whether this is plausible depends on the force of the considerations in favor of the tolerant stance and the force of our reasons to avoid the intransitivities in question. The case in favor of the tolerant stance is simple. The whole-life and future-facing perspectives are both undeniably available to us, and our tendency to manifest future-bias in response to at least some cases of divergence between them is a deep and perhaps ineliminable feature of our agency. Furthermore, most people are strongly inclined to affirm their future-biased preferences upon reflection. Yet, at the same time,
our future-bias is limited in a number of ways. Those limits too are deeply entrenched features of our attitudes, and ones that we are strongly inclined to affirm upon reflection. In both cases, the tolerant stance asserts, the appropriate response to our attitudes is one of acceptance.

What about our reasons to avoid intransitive preferences? As I have said, such preferences are widely taken to be irrational, although there has been some debate about this among philosophers in recent decades. I don’t want to intervene in this general debate, except to the extent of trying to ascertain how far considerations of transitivity cast doubt on the rationality of the bias toward the future. One place to begin is by asking why exactly intransitive preferences are thought to be irrational? If this were simply a stipulation about how the term ‘irrational’ was being used, then it would lack any substantive force. But it is more than that. The claim that intransitive preferences are irrational relies on the thought that if someone prefers A to B, B to C, and C to A, then it strikes us immediately that something has gone substantively awry. If it is not exactly that the person is reasoning badly, then perhaps it is that they don’t really grasp the concept of a preference. The important point is that it is the substantive thought that something has gone awry that underwrites the charge of irrationality.

This thought is often illustrated or reinforced by arguments to the effect that a person with intransitive preferences will be vulnerable to “money-pumping”. That is, such a person will be disposed to accept a sequence of trades that will leave her worse-off than she was when she started. She will, say, pay $1 to move from C to B, $1 to move from B to A, and $1 to move from A to C. So she will end up back where she started, only $3 poorer. And then the cycle can simply start all over again. Clearly, something has gone wrong.

This means that if, by manifesting future-bias in response to some but not all cases of divergence between the whole-life and future-facing perspectives, we are liable to form intransitive preferences in certain situations, then this is a serious obstacle to endorsing the tolerant stance. Before trying to decide whether the obstacle can be overcome, it may help to look at some of the specific sorts of intransitivities that have been said to arise. Philosophers have provided examples of several different kinds, but I will consider just two: one of Hare’s and one of Dougherty’s.

Hare’s example relies on an important observation about the alleged asymmetry, with respect to future-bias, between first-person and third-person cases. As Hare observes, Parfit’s sense that he would experience no future-bias in the case of his distant mother’s illness relies crucially on the fact that she is distant and that he cannot see or communicate with her. In cases where those features are lacking, Hare argues, we do experience a bias toward the future in behalf of other people. Imagine, he says, that you receive a letter on July 28 informing you that your daughter, who is spending the summer at a monastery in Japan, was to have either a more painful tooth extraction on July 27 or a less painful extraction on July 29. Hare thinks that, in this situation, most people would prefer for their daughter to have the less painful extraction on the 29th. Here, as in Parfit’s case, we do not display future-bias in behalf of someone else. But
now Hare asks us to imagine that you receive the letter on July 26th and fly immediately to Japan, arriving at your daughter’s bedside on July 28th. You find her sleeping restlessly, but are uncertain whether she had the more painful operation on the 27th or is scheduled for the less painful operation on the 29th. Her restlessness might be a sign either of post-operative discomfort or of pre-operative anxiety. In this case, Hare believes, most people would prefer for their daughter to have had the more painful extraction on the 27th. If he is right about these cases, then what this shows is that we are prone to experience future-bias in behalf of other people in some contexts but not others. In some contexts, it seems, we give the future-facing perspective of someone else priority over that person’s whole-life perspective. In other contexts, we do not.

This is an important observation in its own right, because it complicates our understanding of, and so helps further to illuminate, the place of these two perspectives in our lives. But Hare also argues that our tendency to toggle back and forth between future-biased and temporally neutral responses in third-person cases, depending on whether the person we are concerned about is either near or distant, may result in the formation of intransitive preferences. He illustrates the point with a variant of the previous cases. Suppose again that you receive the letter on July 26th informing you of your daughter’s situation, but that this time you know only that she is either in a monastery in the north of Japan or in a monastery in the south of Japan. In response, you and your spouse fly off to the two different monasteries. Your spouse, who is better than you at providing post-operative comfort, goes to the southern monastery. You, who are better than your spouse at allaying pre-operative anxiety, head for the northern monastery. You arrive there on July 28th and find a sleeping figure who may or may not be your daughter. What, in this situation, do you hope for? Your daughter
may be nearby or she may be far away. And she may have had the operation already or she may be about to have it. So there are four possibilities:

A. She is nearby and has already had the more painful operation.
B. She is far away and has already had the more painful operation.
C. She is nearby and will have the less painful operation.
D. She is far away and will have the less painful operation.

In this case, Hare says, you will prefer B to A, because if she has had the more painful operation she will be better off with your spouse than with you. And you will prefer C to D, because if she is going to have the less painful operation she will be better off with you than with your spouse. You will prefer A to C, because when she is nearby your future-bias leads you to prefer less pain in her future, and you will prefer D to B, because when she is distant, your neutralist stance leads you to prefer that she experience less pain overall. But this means that you prefer B to A, A to C, C to D, and D to B. Your preferences are intransitive. Notably, the intransitivity depends on the fact that, when comparing A and C, the whole-life and future-facing perspectives diverge. So too for B and D. Yet in preferring A to C, you give the future-facing perspective priority over the whole-life perspective, while in preferring D to B, you do the reverse.

Some people may feel that their own reactions in the cases Hare describes would be either consistently future-biased or consistently neutralist, so that, for them, no intransitivity would arise. Other people may have reservations about some of the details of what Hare says.\footnote{Will you really prefer C to D? Admittedly, your daughter will be better off with you than with your spouse during the pre-operative period, but she will be better off with your spouse during the post-operative period, and if she is going to have her operation tomorrow, she will need to go through both periods. Perhaps we could stipulate that, if} Let’s suppose, however, that many of us would react in just
the ways Hare thinks we would and, accordingly, that we would be led to form intransitive preferences in his final case. Before attempting to assess the significance of this conclusion, let me first describe one of Dougherty’s examples. Dougherty argues that although we display future bias with respect to hedonic experiences of pleasure and pain, we display no such bias in cases where we are contemplating exchanges between hedonic experiences and other goods. Suppose you are willing to exchange twenty dollars now for the gustatory pleasure that a restaurant meal will bring you tomorrow. So you think this future pleasure is worth twenty dollars. If you were relevantly future-biased, however, you might well think, after having eaten the meal, that you had overpaid, because the meal and its associated gustatory pleasure would then be in the past, and a past meal, with its associated pleasure, would not be worth as much money to you as a future meal, with its associated pleasure. So, if you were relevantly future-biased, you would regret having paid twenty dollars for the meal. After all, if the twenty dollars were still available to you, you could use it to purchase future gustatory pleasure, and you prefer future gustatory pleasure to past gustatory pleasure. Yet from a whole-life perspective, there is no basis for such a preference and, accordingly, no basis for regretting your purchase of the past meal. From that perspective, it makes no difference whether your gustatory pleasure lies in the future or the past. And in fact, Dougherty says, you probably won’t regret your purchase. With respect to exchanges between hedonic and non-hedonic goods, he claims, we generally see things from a whole-life perspective and remain temporally neutral. Or, as he puts

C obtained, your spouse would be able to travel from the southern monastery to the northern monastery in time to provide your daughter with post-operative comfort. So your daughter would have the less painful operation and she would receive optimal pre-operative and post-operative care. If so, however, then is it still clear that you would prefer A to C?
it, we maintain a “fixed exchange rate” between hedonic and non-hedonic goods, and we think it rational to do so. But if we have fixed exchange rates in such cases, then we cannot, on pain of violating the global requirement to have transitive preferences, be future-biased in cases involving only hedonic goods. For if, given my fixed exchange rate, I am indifferent between a dollar and a minute of future gustatory pleasure, and equally indifferent between a dollar and a minute of past gustatory pleasure, then transitivity requires that I also be indifferent between a minute of past pleasure and a minute of future pleasure. If, like most of us, I prefer future pleasure to past pleasure, then my preferences are not transitive.

Again, we may or may not agree with what Dougherty says about this example. For the sake of argument, though, let’s suppose that both he and Hare have provided examples in which people are likely to alternate between future-biased and neutralist responses and, in consequence, to form intransitive preferences. In Hare’s example, the intransitivity results from the fact that, in our attitudes toward our loved ones’ sufferings, we seem to toggle back and forth between neutralist and future-biased preferences depending on whether our loved ones are nearby or far away. In Dougherty’s example, the intransitivity results from the fact that, in our attitudes toward exchanges of goods, we seem to toggle back and forth between neutralist and future-biased responses depending on whether the exchanges are between hedonic goods only or between hedonic and other kinds of goods. How troubled should we be by these intransitivities? How great an obstacle do they present to endorsement of the tolerant view of future-bias?

On one interpretation, the point of the examples is simply to illustrate how intransitive preferences can arise in cases involving future bias. It is not to establish the irrationality of those preferences, which is instead taken for granted. So interpreted,
however, these examples won’t help us very much if what we are trying to do is precisely to decide whether the kind of intransitivity to which we may be liable if we are sometimes but not always future-biased is a manifestation of irrationality. We have already agreed that, if someone has intransitive preferences, then it seems to most people as if something has gone seriously awry. Yet the examples depend for their force on the thought that, although future bias seems appropriate to most people in some cases of divergence between the whole-life and future-facing perspectives, temporal neutrality seems appropriate in others. This means that, if we imagine trying either to eliminate future-bias altogether or to extend its scope more widely, we are likely to meet with internal resistance: not only with a feeling that it would be difficult to enforce such a decision, but also with a sense that something would have gone awry if we did. It may look, then, like we are at an impasse, and we cannot overcome the impasse by simply assuming that intransitivity is always irrational.

Sometimes when people give examples in which a certain pattern of concern is said to lead to the formation of intransitive preferences, a different interpretation of these examples may be intended. The point may be to establish that we are liable to engage in self-defeating behavior if we exhibit the target pattern of concern and, accordingly, that our intransitive preferences are problematic not only theoretically but also practically. We may, for example, be vulnerable to the kind of exploitation that is involved in money-pumping. On this interpretation, the point of such examples is not merely to illustrate the irrationality of intransitive preferences; it is to provide an independent argument for it. But in the examples of Hare’s and Dougherty’s that I have described, there is no possibility of money-pumping, because the options among which one is forming one’s preferences include different past histories, and there is no way that a money-pumper could credibly offer to bring about the past history of one’s
choice. Even if we were vulnerable to money-pumping in those cases, moreover, the
claim that our intransitive preferences were irrational for that reason would take us
back to the sort of argument that Parfit rejected, the argument that says future-bias is
irrational because it is bad for us. And even if we were prepared to admit such
arguments as legitimate in principle, the badness of the specific intransitivities arising
from future-bias (in combination with certain other attitudes) would have to be judged
comparatively. We would have to decide whether the practical problems exemplified
by the threat of the money pump in those cases were more severe than whatever
practical disruptions would be produced by the elimination of future bias. This would
require not only an assessment of the severity of the money-pumping threat but also a
comprehensive accounting of the advantages and disadvantages of Timelessness. And,
as I have argued, we do not yet have such an accounting. Furthermore, even if, in
addition to making the suppositions already enumerated, we also supposed that the
money-pumping threat was the greater practical problem, that would still not show that
future bias was irrational: only that, in the relevant cases, we should display either more
or less future-bias than most of us are now disposed to do.

Dougherty thinks that we tend to overlook the intransitivities to which he calls
attention because our thinking is “compartmentalized.” When we think about
exchanges between hedonic goods and goods of other kinds, we conceive of ourselves
as temporally extended creatures and take a neutralist stance. But when we think about
exchanges solely among hedonic goods, we view things perspectivally and exhibit a
bias toward the future. Our tendency to treat the two sorts of exchange differently

---

In “On Whether to Prefer Pain to Pass” (Ethics 121[2011]: 521-537), Dougherty gives a
different example which is meant to show that people who are both future-biased and
risk-averse can be turned into “pain pumps.”
derivates from these “twin aspects of our identities,” and our compartmentalized thinking leads us to overlook the “tension” between them. By calling attention to the intransitivities that result when exchanges of the two different kinds are considered as part of a single set of options, Dougherty aims to “expose this compartmentalization,” and to help us overcome it. He wants us to recognize that “fundamentally we are temporally extended agents,” and, accordingly, that we should always give priority to what I have called the whole-life perspective. It is this perspective that reflects “who we really are,” and by giving it priority over the future-facing perspective whenever conflicts arise we can overcome compartmentalization and avoid intransitivity.

But whether compartmentalization is a bug or a feature may be in the eye of the philosophical beholder. Some will be tempted to turn Dougherty’s diagnosis on its head. Since the whole-life and future-facing perspectives are both available to us, and since, in that sense, both of them are part of “who we really are,” we cannot expect to eliminate either of them and compartmentalization may be a reasonably effective technique for managing the tension between them. If, by failing to apply the norm of transitivity across “compartments” in the same way that we do within compartments, we accommodate both perspectives while minimizing agential dissonance, and if, in this way, we manage to stay out of practical trouble too, then it is not clear that this is irrational.

---

33 “Future-Bias and Practical Reason,” p. 15.
34 Ibid., p. 13.
35 Ibid., p. 15.
36 Ibid.
9. Future-Bias and Regret-Avoidance

Our tendency to compartmentalize is also relevant to an argument against future-bias that has been developed by Preston Greene and Meghan Sullivan. Unlike Hare and Dougherty, Greene and Sullivan do not focus on cases of cycling or intransitivity. In a somewhat similar spirit, however, they maintain that future-bias in combination with certain other attitudes can lead to self-defeating behavior. On that basis, they argue that the bias is irrational. Simplifying a bit, their main argument is roughly as follows. It is rationally permissible to avoid acting in ways one is certain one will later regret, where regret is understood not as an affective state but rather as a preference that one had acted otherwise. But future-biased agents who wish to avoid certain regret will have to postpone their pleasurable experiences for as long as possible, if they are to avoid regret about those experiences being in the past. If Jack chooses to have an enjoyable meal on Monday rather than waiting until Friday, when those are the options available to him, then as soon as the meal is over, he will strongly prefer that it were in the future rather than the past, which is to say he will regret having scheduled it for Monday. And if future-bias is rational, then Jack’s regret is rational. Greene and Sullivan call this the scheduling problem. Even worse, future-biased agents who wish to avoid certain regret will have to forego greater goods in the immediate future in order to secure lesser goods in the further future. Billy might forgo, say, two cookies now in order to be assured of having one cookie, or even a morsel of a cookie, at some point in the future. After all, if he chose to have two cookies now, then as soon as he had eaten them he would regret his choice, since his eating

In “Against Time Bias.”
experience would then be in the past, and he could instead have had a cookie-eating experience still ahead of him. Greene and Sullivan call this the meager returns problem.

According to Greene and Sullivan, it is irrational to act in ways that lead to the scheduling and meager returns problems. Yet if one chooses to avoid certain regret and one is future-biased, one will act in these ways. So either it is not permissible to avoid certain regret or future-bias is irrational. But the former, they say, is implausible; surely there is no rational requirement that one not avoid acting in ways that one will certainly regret. It follows that future-bias is irrational.

One thing that is striking is that, although we are future-biased, most of us do not act in ways that lead to the scheduling and meager returns problems. We do not postpone restaurant meals indefinitely and we don’t exchange present cookies for future cookie-morsels. How is this to be explained? One suggestion is that, in general, we simply don’t care about avoiding certain regret. But this does not seem true. Although we sometimes do things we know we will regret, usually we try to avoid doing that. Another suggestion is that our bias toward the near offsets our bias toward the future, so the combination of the two biases prevents us from acting irrationally in the ways Greene and Sullivan describe. For example, our impulse to accept a cookie morsel later in preference to two whole cookies sooner, and thereby to avoid future regret, is offset by our near-biased impulse to prefer cookie pleasure sooner to cookie pleasure later. But if that were the whole story, then we would expect people routinely to experience regret once their restaurant meals were over and their cookies consumed, since at that point their bias toward the near would no longer be engaged and their bias toward the future would lead them to prefer that they still had the relevant pleasures

*Greene and Sullivan themselves appear to endorse this suggestion, at least as it applies to the scheduling problem. See “Against Time Bias,” p. 960.*
ahead of them. Yet it is not clear that most people are subject to all-things-considered preferences of this kind. If not, then perhaps the reason that future-biased people who wish to avoid regret nevertheless do not fall prey to the scheduling and meager returns problems is that they do not actually expect to experience regret in the relevant cases.

Why wouldn’t they experience such regret? Presumably because their future bias does not come into play, or does not come into play with full force, in these cases. But why not? It may be relevant that, in the cases that Greene and Sullivan present, the putatively regrettable outcome comes about, if it does, as a result of one’s own choice, rather than resulting from factors exogenous to one’s choices. This contrasts with Parfit’s hospital case, for whether you had the longer operation yesterday or will have the shorter operation today is not a function of any choice you did or will make. But one chooses to have two cookies now rather than a cookie morsel later. At the moment when one makes that choice, moreover, the two-cookie alternative is (possible regrets aside) preferable from both the future-facing and whole-life perspectives. In other words, both one’s future and one’s life as a whole will go better if one has two cookies now rather than a cookie morsel later. And perhaps our future-bias does not come into play, at least not with full force, in cases with this structure. Perhaps, more precisely, we do not manifest future bias in cases where 1) the putatively regrettable outcome is one that we ourselves chose, and 2) we chose that outcome, at least in part, because it was (as good as or) better than the available alternatives from both the future-facing and whole-life perspectives. This may also help to explain why, in Dougherty’s case, we are unlikely to feel regret after eating and paying twenty dollars for a meal we had judged in advance to have been worth that much. If these suggestions are correct, then the reason we don’t incur the scheduling and meager returns problems, despite being future biased, is that we don’t expect our future-bias to lead to regret in the relevant
range of cases, and so there is no regret to be avoided. If that is the explanation, it reveals another limit on the scope of our future bias.

But perhaps the fact that we don’t act in ways that lead to the scheduling and meager returns problems should be explained in a different way. Perhaps, even if we do expect to experience regret in cases of the kind described by Greene and Sullivan, we see no need to avoid regret in cases of that type. Does this mean that we deny the permissibility of avoiding regret in such cases? That is not what I am suggesting. Greene and Sullivan acknowledge that, rather than flatly denying the permissibility of avoiding certain regret, one might instead hold that “one can [permissibly] aim to avoid regret except in cases in which doing so leads to the scheduling or meager returns problems.” They allow that this modified norm would avoid the conclusion that future-bias is irrational, but they also say that it would be unacceptably ad hoc. My suggestion, however, is not that we affirm such a norm, but rather that we simply don’t in practice seek to avoid certain regret in these cases. And while one might say that if this is so, then we are guilty of drawing ad hoc distinctions, one might instead say that it reveals yet another way in which our thinking is compartmentalized. Of course, a critic might insist that such compartmentalization is itself irrational. But perhaps the correct conclusion is that, for creatures with our repertoire of attitudes, it is a reasonably effective way of avoiding irrationality and staying out of practical trouble.

In short, I have suggested two possible explanations for why we don’t generally act in ways that lead to the scheduling and meager returns problems. One possibility is that, owing to an implicit limit on the scope of our future bias, there is no future regret to be avoided in the relevant cases. The other possibility is that, owing to our tendency

---

*“Against Time Bias,”* p. 965.
to compartmentalize, we simply don’t treat future regret as something to be avoided in those cases. The two explanations might be combined; perhaps, up to a point, we don’t expect to experience regret in the relevant cases and, beyond that point, we don’t treat the regret as something to be avoided. Consider again the cookie example. Perhaps, on the one hand, we don’t expect to experience much regret about having chosen the two cookies after we have eaten them, even though we will then have no cookie morsel to look forward to; and perhaps, on the other hand, even if we do anticipate some regret, we don’t see that regret as something to be avoided. Whatever the explanation, the fact remains that the combination of future bias and a desire to avoid regret does not in practice lead us to act irrationally. So it is not clear why reflection on that combination of attitudes should lead us to condemn our future-bias as irrational.

10. Conclusion

It is time to draw this discussion to a close. I have not considered all the arguments that have been put forward against the rationality of future-bias. “For that reason and others, what I have said has not been conclusive. But I hope I have managed to explain why I am not yet ready to abandon the tolerant stance or to join Sidgwick, Rawls, Parfit and the many others who have advocated temporal neutrality and who have, either explicitly or implicitly, condemned the bias toward the future as irrational. For now, at least, I am inclined to align myself instead with McTaggart, who

“For a sustained defense of temporal neutrality, see Meghan Sullivan, Time Biases (Oxford University Press, 2018). Many but not all of Sullivan’s arguments against future-bias in that book recapitulate arguments I have addressed in this essay, including arguments of Dougherty’s that I have discussed as well arguments advanced by Greene and Sullivan in “Against Time Bias.”
writes: “the anticipation of a good in the future produces greater happiness in the present than the memory of an equal good in the past. There may be no more reason for this than there is for a man’s preference of burgundy to claret, or claret to burgundy. But the absence of reason is not here contrariety to reason. It would, no doubt, be possible for a man to hold that his nature would be more admirable if good and evil affected him to the same extent when they were past as when they were future. But I do not know any reason why anyone should hold this.”