

## Time – The Emotional Asymmetry

### **1. Future-Bias**

A person is *future-biased* when she would rather, other things being equal, that bad things be in the past than in the future, and that good things be the future than in the past.

Most of us are, at least to some degree, future-biased. Consider:

#### Your Past or Future Pain

You wake up with an aching jaw. What is going on? You remember that you were scheduled to have your wisdom teeth removed, painfully but safely, under a weak local anesthetic, on Thursday. But has that happened yet? In your groggy condition you are not sure. It could be Friday morning. The ache in your jaw would then be an after-effect of the operation. And it could be Thursday morning. The ache in your jaw would then be the distress of your impacted teeth.

I expect that in this situation you would want it to be Friday. You would want your pain to be in the past.

Indeed, for most of us, our future-bias is strong enough that sometimes we would rather that bad things be in the past than in the future when other things are not equal. Sometimes we would rather that worse things be in the past than that less bad things be in the future. Consider:

### Your Greater-Past or Lesser-Future Pain

Again you wake up with an aching jaw. What is going on? This time you know that it is Thursday morning, but you cannot quite remember what was scheduled to happen to you. Either you were to have your wisdom teeth removed under a weaker anesthetic on Wednesday, or you were to have them removed under a stronger anesthetic on Thursday. Not to worry too much about this, though. Both kinds of operation are safe, and their after-effects are exactly the same.

I expect that in this situation you would rather that you have had the Wednesday operation than that you will have the Thursday operation. You would rather that you experienced more pain, yesterday, than that you will experience less bad pain, today.

I will address two questions concerning future-bias here. First, with respect to what sorts of things are we future-biased? Is it that we want *all* things that we regard as bad to be in the past, or *just some* of them? Second, are we justified in being future-biased with respect to these things? This second question has received a good deal of attention from philosophers (a wave of attention followed some enigmatic remarks by Arthur Prior (in Prior 1954), another wave followed an extended treatment of the topic by Derek Parfit (in Chapter 8 of Parfit 1984)). My aim here is to survey different answers to the question, and to give a sense of how things presently lie.

## 2. With Respect to What Sort of Things are We Future-Biased?

This is a question that psychologists are best-placed to answer. But, as far as I know, psychologists have paid relatively<sup>i</sup> little direct attention to it,<sup>ii</sup> so let's make do with armchair resources here. I am not a psychologist, but I possess a moderately normal psychology, and I can say things based on informal observation of it.

First, I am future-biased, not only with respect to my pain, but also with respect to other experiences of mine that I take to be bad, experiences that are not so naturally described as 'pain'. Consider:

### My Past or Future Awkwardness

I wake up wincing. What is going on? I remember that I was scheduled to have an awkward meeting with a colleague on Thursday – a meeting sure to elicit disappointment, anger and embarrassment in us both. But has that happened yet? In my groggy condition I am not sure. It could be Friday morning. It could be Thursday morning.

I guess that in this situation I would want my awkwardness to be in the past. Indeed the desire would be as acute, as pressing, as my desire that dental pain be in the past in Past or Future Pain.

Second, I am not future-biased with respect to things that I take to be bad for me when the badness-for-me does not have to do with my having bad experiences.

Consider:

### Past or Future Infidelity

I learn that my wife plans to be unfaithful to me, for the first time, this week. I very much do not want this to happen. But, being a non-

confrontational sort of a fellow, I decide not to try to prevent it from happening. I isolate myself from her – I retreat to Yorkshire and work on some philosophy. Time passes. Has the dread event occurred? I don't know.

I guess that in this situation I would not in any way care about whether her infidelity was in the immediate past or the immediate future.

Third, I am not future-biased with respect to things that I take to be bad, though not particularly bad for me, when the badness does not have to do with anybody's having bad experiences. Consider:

#### The Jets' Past or Future Victory

I very much do not want the New York Jets to win today's Superbowl. I regard their success as a very bad thing. But they are heavy favorites. I isolate myself from the game – plug my ears, unplug my television, and work on some philosophy. Time passes. Has the dread event occurred? I don't know.

I guess that in this situation I would not in any way care about whether the Jets' victory was in the immediate past or immediate future.

Fourth, and last, my attitude towards other people's bad experiences is curiously dis-uniform. Sometimes I am future-biased with respect to other people's bad experiences. Consider:

#### My Nearby Daughter's Greater-Past or Lesser-Future Pain

I learn that my daughter, away at college, was scheduled either to have her wisdom teeth removed under a weaker anesthetic on Wednesday, or

to have her wisdom teeth removed under a stronger anesthetic on Thursday (not to worry too much about this, though – both operations are safe and their after-effects are exactly the same). I drive over to see her on Thursday morning. I find her restlessly sleeping. Has she had the operation yet? I don't know.

I guess that, in this situation, I would rather that my daughter have had the operation on Wednesday. I would rather that she experienced worse pain, yesterday, than that she will experience less bad pain, today.

But sometimes I am not future-biased with respect to other people's bad experiences. Consider:

My Distant Daughter's Greater-Past or Lesser-Future Pain

I learn, by letter, that my daughter, away at a distant, monastic retreat, far from phones or email, was scheduled either to have her wisdom teeth removed under a weaker local anesthetic on Wednesday, or to have her wisdom teeth removed under a stronger local anesthetic on Thursday. In either case, I will have no contact with her for another two months. It is Thursday morning. Has she had the operation yet? I don't know.

I guess that in this situation I would rather that my daughter have the operation on Thursday. I would rather that she will experience less bad pain, today, than that she have experienced worse pain, yesterday.

What cues me to take different attitudes in the two cases? Why do I care, in the former case, about how much pain she will experience in the future, and in the latter case about how much pain she will have experienced over the course of her

life? *Distance* has something to do with it. In the former case she is near, in the latter case far. *The psychological salience of her present condition* has something to do with it. In the former case I am vividly aware of her as she is now, but in the latter case I am not. *Communication* has something to do with it. In the former case I will be in a position to communicate with her soon, at a time when I know that she (being future-biased on her own behalf) will wish that she had the more painful, Wednesday operation, but in the latter case I will not be in position to communicate with her until much later, at a time when I know we will both wish that she had the less painful, Thursday operation.

These seem to me to be the factors that trigger future-bias on behalf of my daughter. Whether they *should* trigger future-bias on behalf of my daughter is another matter. I will address it in section 5.

To summarize this section: I am future-biased with respect to my own bad experiences. I am not future-biased with respect to things that I take to be bad (whether bad for me, like my wife's infidelity, or bad in some other way, like the Jets' victory) where the badness does not have to do with bad experiences. I am selectively future-biased with respect other people's bad experiences. And I believe I am not abnormal in these respects. Other people tell me they feel the same way.

### **3. Some Poor Arguments for Future-Bias**

Future-bias is deeply engrained in us. It is there before we study physics or philosophy, before we have any educated views about the nature of time, about what it is for an event to be past or future. But physicists and philosophers have

thought hard and learned a great deal about the nature of time. In light of what they have learned, does it make sense to be future-biased in the way we are future-biased?

An analogous question (as a warm-up): We tend to be *mammal-biased*. We tend to care, for example, more about the pain of ferrets than about the pain of mackerel. That pattern of concern is there before we have any educated views about animal physiology. But scientists have learned a great deal about animal physiology. In light what they have learned, does it make sense to be mammal-biased?

There are three basic positions to take. We could say that we ought to be future-biased, that we would be making some kind of mistake in failing to be so. We could say that we ought not to be future-biased, that we are making some kind of mistake by being so. And we could say that it is fine either way.

Let's start with the first position. We ought to be future-biased. Why? When I ask people this question, I find that some, very carefully, answer:

"Past pains should matter less to us because they are past, future pains should matter more to us because they are future. At some point explanations and justifications must stop. This is one of those points."

And many more answer, less carefully, in what amounts to the same way. They say things like:

"Because past pains have happened, future pains are yet to happen."

"Because past pains are gone forever, future pains are still to come."

"Because the past is behind us, the future ahead of us."

But it is not clear what they achieve, by drawing attention to the fact that past pains 'have happened', or that they are 'gone forever', or that they are 'behind us', beyond drawing attention to the fact that that past pains are past.

Maybe this sort of answer is right. But we should not settle for it without first considering whether there is a good, persuasive argument to the conclusion that we ought to be future-biased – an argument whose premises will be accepted by people who are not antecedently committed to the conclusion.

One thing such an argument could do is point to some difference between past and future pains, beyond that the ones are past and the others future, in virtue of which it makes sense for us to care more about the future pains.

What sort of differences could play this role? A candidate difference has to do with *existence*. Suppose the *shrinking block theory* of time is correct – future events exist, past events do not. Then perhaps it would make sense to care more about future pains than past ones, because it makes sense to care more about what is than about what is not.

But this is not a very satisfactory argument. First, note that to make sense of the shrinking block theory of time we must use the term 'exist' in such a way that it is not an analytic truth that all things that exist exist-in-the-present-moment. But once it is clear that we are using the term this way (to mark that they are using the term this way, philosophers often speak of 'tenseless existence' – I will follow suit here) it is perhaps not so obvious that it makes sense to care more about what tenselessly exists than about what does not tenselessly exist. Furthermore the shrinking block theory of time has it that my past pains tenselessly existed – the



block *was* larger than it now. And it is perhaps not so obvious that it makes sense to care more about what tenselessly exists than about what tenselessly existed.

Second, if the shrinking block theory did vindicate future-bias with respect to our own bad experiences then it would seem to vindicate future-bias with respect to all bad things. But we are not future-biased with respect to all bad things (think of my wife's infidelity, my distant daughter's pain, the Jets' victory). So it will not vindicate our entire package of future-biased attitudes. Third, I know of no independent reason to accept the shrinking block theory of time – nothing that would move a future-unbiased person to accept it.<sup>iii</sup>

Another candidate difference has to do with *metaphysical openness*. Maybe the future is metaphysically open, the past not. One way (of several) of making sense of this idea: Say that an event is *presently settled* when either it is the case that the event is occurring or will occur or has occurred, or it is the case that the event is not occurring and will not occur and has not occurred. Maybe the norm is for past events to be presently settled but future events not. So Nelson's victory at the Battle of Trafalgar is presently settled, but Obama's victory in the election of 2012 is not – it is neither the case that it will occur nor the case that it will not occur. Then it would make sense to care more about future events than past ones, because it makes sense to care more about what is presently unsettled than what is presently settled.

But this is not a very satisfactory argument either. Perhaps it does make sense for us to direct our attention towards presently unsettled events, but it does not follow that it makes sense for us to prefer that episodes of pain be presently

settled rather than presently unsettled. The opposite seems true. Why not prefer that the bad event be presently unsettled?

Yet another candidate difference has to do with *epistemic openness*. We have very different epistemic access to the past and future. We learn about the past through memory, through record, through reconstruction. We learn about the future through anticipation, through prediction. The latter modes of access are less robust than the former. We tend to know more about the past than the future. Could this somehow vindicate future-bias? It does not look promising. Why would it make sense to want pain to be such that I am in a better position to know of it?

Yet another candidate difference has to do with *causation*. We often cause future events to occur, but rarely, outside of science fiction, cause past events to occur. Could this somehow vindicate future-bias? Again, it does not look promising. Why would it make sense to prefer that I have no control over whether a pain event occurs than that I have control over whether a pain event occurs? And, in any case, we very often do not have control over future events. We can readily tweak the Past or Future Pain examples so that, no matter whether the pain-event is past or future, I have no control over whether it occurs.

But perhaps there is a different sort of argument that turns on our ability to control aspects of the future. Here's a simple evolutionary story (for more sophisticated stories see Maclaurin and Dyke 2002 and Suhler and Callender forthcoming): It is not an accident that we are future-biased with respect to pain. That feature of ourselves has been selected-for by evolution. In light of the direction of causation, ancestral creatures that focused their practical attention on the future

did better than their peers that focused their attention on the past. And a cognitively efficient way to focus a creature's practical attention on the future is to have the creature care a great deal about its future pains and not at all about its past pains – a pattern of concern that quite naturally yields a preference for pain being past rather than future.

Maybe the same feature that served our swampy ancestors well in their quest to pass on DNA serves us well in our quest to have good lives. We are better off for caring a great deal about our future pains and not at all about our past pains.

This argument, if it worked, would give us what is known as a *state-based* justification for future-bias. It does not address the desirability of our pain being past rather than future. It does not point out a difference between past and future pains, in light of which it makes sense for us to prefer that pain be past. It addresses the desirability of our preferring that pain be past. It points out a difference between preferring that pain be past and failing to do so, in light of which it makes sense for us to desire that we prefer that pain be past. State-based justifications for mental attitudes have notoriously limited persuasive force (because the move from your taking something to be desirable to your desiring it is seamless, while the move from your taking it to be desirable that you desire something to your desiring it is much less so.) But still, perhaps we would have made some progress towards justifying future-bias.

But the argument does not work. Whatever the capabilities of my swampy ancestors, I think that I am well capable of focusing my practical attention on future things over which I have control without being future-biased. And any loss of

cognitive efficiency that would accompany this has to be balanced against the benefits that would accompany it. As Parfit himself pointed out (in Chapter 8 of Parfit 1984), when you are old and your future is short and dusky it may well be in your interest to be future-unbiased.

That is, of course, a contingent feature of me and my circumstances. Maybe for some people, in some circumstances, it will turn out to be beneficial to be future-biased. But the same holds for all practical attitudes.

Will any other argument work? I know of no good, persuasive argument for future-bias. If I were to come across future-unbiased, but otherwise rational alien, then I will not suppose myself able to convert her by force of reasoning. More personally: if you lose your future-bias then reasoning will not recover it for you. Not to worry, though, because you won't lose it – you may as well try to lose your skin.

#### **4. The Arbitrariness Argument Against Future-Bias**

The opposite position is that we are making a mistake in being future-biased? We should be future-unbiased. Why? One reply:

“It just does not matter whether a pain is past or future. We have no reason to care about that. So future-bias is objectionably *arbitrary* – in that it involves caring about things that we have no reason to care about.”

Again, this may be right. But again, it does not give us a good, *persuasive* argument to the conclusion that we should not be future-biased. Only the

antecedently persuaded will accept its major premise – that we have no reason to care about whether pain is past or future.

One way to get such an argument would be to give an account of what it is for pain to be past or future that would incline the antecedently unpersuaded to say

“If *that* is what it is for pains to be past and future, well, clearly we have no reason to care to care about *that*.”

There are hints of such an argument in A.N. Prior’s 1954 article “Thank Goodness That’s Over”. Prior criticized philosophers (notably Quine) who were engaged in the early stages of the project of giving a semantics for tensed utterances (like ‘it is *now* raining’, ‘it *was* sunny yesterday’, ‘it *will be* sunny tomorrow’) using a tenseless metalanguage (a metalanguage containing only expressions like ‘rain on Monday precedes sun on Tuesday’). Prior wrote:

One says, e.g. ‘Thank goodness that’s over!’, and not only is this, when said, quite clear without any date appended, but it says something which it is impossible that any use of a tenseless copula with a date should convey. It certainly doesn’t mean the same as, e.g. ‘Thank goodness the date of the conclusion of that thing is Friday June 15 1954’, even if it be said then. (Nor, for that matter, does it mean ‘Thank goodness the conclusion of the thing is contemporaneous with this utterance’. Why should anyone thank goodness for that?)

This passage is famous, in large part because the project of giving a tenseless semantics for tensed utterances has endured, motivated by metaphysical considerations. The world is a four-dimensionally extended space-time manifold.

Things within the world may have relational tensed properties, but they do not have further, non-relational tensed properties. So the events of 2002 have the relational tensed properties of *being future relative to the events of 1992* and *being past relative to the events 2012*, but they do not have the further, non-relational tensed property of *being past*. Since relational tensed properties can be described in tenseless language (e.g. 'the events of 2002 precede the events of 2012'), we can describe all properties of things in tenseless language. So, insofar as tensed utterances are made true by things having properties, we can describe what makes them true in tenseless language.

Many philosophers have understood Prior to be posing a kind of challenge to the four-dimensionalist: What is that Prior is thanking goodness for, when he thanks goodness that his troubles are over? In general, what is your account of the *content* of tensed propositional attitudes? This challenge has been answered in different ways. Perhaps the most influential answer, due to David Lewis (1979), has it that wanting pain to be over involves wanting of myself (construed as a person-stage – a thing that does not persist over time) that I be to-the-future of pain.

But there is a different way of understanding Prior. If the four-dimensionalist metaphysics and semantics is correct then his attitude of thanking goodness that his troubles are over does not make sense. Why should I want my pains to be in the past if wanting pain to be in the past just amounts to wanting of myself (construed as a person-stage) that I be to the future of pain? Wouldn't that be just like my standing in a row of soldiers at a parade ground, knowing that one of them has a toothache,

and wanting of myself that I be to the south of the pain? And wouldn't such a desire be objectionably arbitrary?

Now it must be said that, though the universe may be a four-dimensionally extended manifold, the temporal dimension is not just like the spatial dimensions. There are interesting asymmetries along the temporal dimension that do not exist along the spatial dimensions – epistemic asymmetries, causal asymmetries, entropic asymmetries. But, as we saw in the last section, it is hard to see why any of these things give me a reason to be future-biased. So, by pointing to these things, we do not explain why a desire that I (construed as a person-stage) be to the-future-of pain is less arbitrary than a desire that I be to the south of pain.

It also must be said that, even if four-dimensionalism is false, a charge of arbitrariness can be leveled at future-bias. One way for four-dimensionalism to be false is for *presentism* to be true. My past and future pains do not tenselessly exist. Tensed locutions like 'it was that' and 'it will be that' should be understood as sentential operators for which no reductive semantics can be given. My wanting my pain to be past involves my preferring that *it was that* (I am in pain) rather than that *it will be that* (I am in pain). But why care about that? Another way for four-dimensionalism to be false is for *tense realism* to be true. My past and future pains tenselessly exist, but their being past or future consists in their having non-relational tensed properties. My wanting my pain to be past involves my preferring that it have the non-relational property *being past*, rather than the non-relational property *being future*. But why care about that?

Prior's implicit idea (the way we are reading him here) is that we face a special burden of explanation when it comes to attributing significance to the way things stand in relation to ourselves. It is entirely familiar and unproblematic that we take it to matter whether pain has certain intrinsic features, but can give no persuasive argument as to why. I think it matters whether pain is *less intense* or more *intense*. I do not think that I could persuade an intensity-unbiased but otherwise rational alien of this. But this does not, and should not, bother me. It is also entirely familiar and unproblematic that we take it to matter whether certain sentences containing sentential operators for which no reductive semantics can be given are true, but can give no persuasive argument as to why. I think it matters whether *it is the case that* (I am in pain) or *it might have been, but is not the case that* (I am in pain). I do not think that I could persuade an actuality-unbiased but otherwise rational alien of this. But this does not, and should not, bother me. But it is not okay to take it to matter whether I stand in a certain relation to pain, but be unable to give a persuasive argument as to why.<sup>iv</sup>

I am quite taken by this idea (see Hare 2009). But for present purposes that does not matter. What matters is that there is some distance between accepting the idea and rejecting future-bias. Prior took himself to be arguing against four-dimensionalism, not future-bias. The idea tells against future-bias only if we have a strong antecedent commitment to four-dimensionalism. If we have no such commitment then we can remain untroubled.



## 5. Incoherence Arguments Against Future-Bias

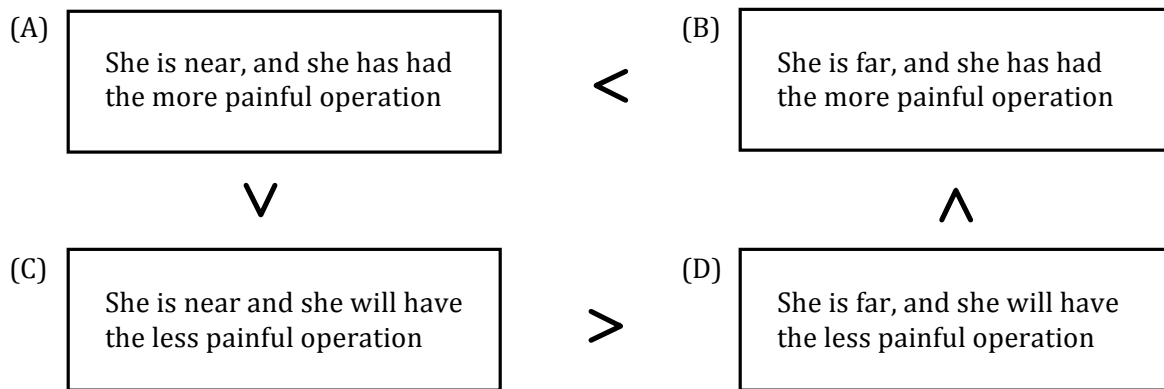
Some patterns of future-biased attitudes may exhibit a kind of structural incoherence. Look back at my attitudes towards my daughter's pain. I said that under some conditions (when her present condition is salient to me, when she is right there before me) I am future-biased with respect to her pain, but under other conditions (when her present condition is not salient to me, when she is far away, when I know that I will not interact with her for a long time) I am not future-biased with respect to her pain. But, whether she is near or far, I will always prefer, other things being equal, that she experience less pain in the future and less pain over the course of her life than more pain in the future and more pain over the course of her life. Now consider (I am here repeating an argument from Hare 2008):

### My Nearby or Distant Daughter's Greater-Past or Lesser-Future Pain

Once again I learn, by letter, that my daughter, away at a distant, monastic retreat, far from phones or email, was scheduled either to have her wisdom teeth removed under a weaker local anesthetic on Wednesday, or to have her wisdom teeth removed under a stronger local anesthetic on Thursday. But this time I am unsure where the letter came from. She is either staying in a monastery in the far-north of Japan or staying in a monastery in the far-south of Japan. Undeterred, my wife and I jump on planes – hers heading to the south, mine heading to the north. On early Thursday morning I arrive at the northern monastery and am confronted with a sleeping figure. In the dim light I cannot quite tell if it is my daughter, and certainly cannot tell if my daughter has had her operation yet. One thing I do know is this: I am significantly better than my wife at allaying pre-operative anxieties (with games, funny

stories...etc.) and my wife is significantly better than me at providing post-operative comfort (with hugs, grapes... etc.).

How might things be? She might be near or far. She might have had the operation or be about to have the operation. How do I want things to be? Well, of the four possibilities:



I prefer (B) to (A) – because, if she has had the more painful operation then she will have a better future, and no better or worse a past, if she is far away, with my wife. I prefer (C) to (D) – because, if she is will have the less painful operation then she will have a better future and no better or worse a past, if she is right here with me. I prefer (A) to (C) – because, if she is near then I would rather that she have the better future. I prefer (D) to (B) – because, if she is far then I would rather that she suffer less pain over the course of her life. My preferences are cyclical.

On the assumption that rational people do not have cyclical preferences, this shows that I cannot, if I am rational, be future-biased with respect to my daughter’s pain when she is near, future-unbiased with respect to her pain when she is far, and

always prefer that she suffer less pain in future and less pain over the course of her life, rather than more pain in future and more pain over the course of her life, no matter whether she is near or far. Something has to give.

But, of course, it does not show that there is something quite general wrong with future-bias. I can avoid incoherency just by becoming less selective in my future-bias towards my daughter. Tom Dougherty has recently (Dougherty 2011) presented a very interesting coherence argument to the stronger conclusion that there is indeed something quite general wrong with future-bias.

All of the examples we have seen so far involve people awaiting news. Future-bias plays a role in determining their reactions to the news (e.g. in determining whether they are relieved upon hearing that they had the more painful operation yesterday) but it does not play any role in determining how they act. And one might think that this is quite generally true. Future-bias is not action-guiding in realistic contexts. The characteristic preferences are between states of affairs with different past components (e.g. a state of affairs in which you suffered more pain yesterday and a state of affairs in which you suffered less pain yesterday). But, outside of science fiction, we are never in a position to bring about states of affairs with different past components. Outside of science fiction, we cannot change the past.

Dougherty's first observation is that this is not quite right. Sometimes our actions may be guided, not by a desire to change the past, but by our views about whether it is good or bad that the past be one way or another. If you are risk averse, for example, then you will act so as to protect yourself against the possibility of

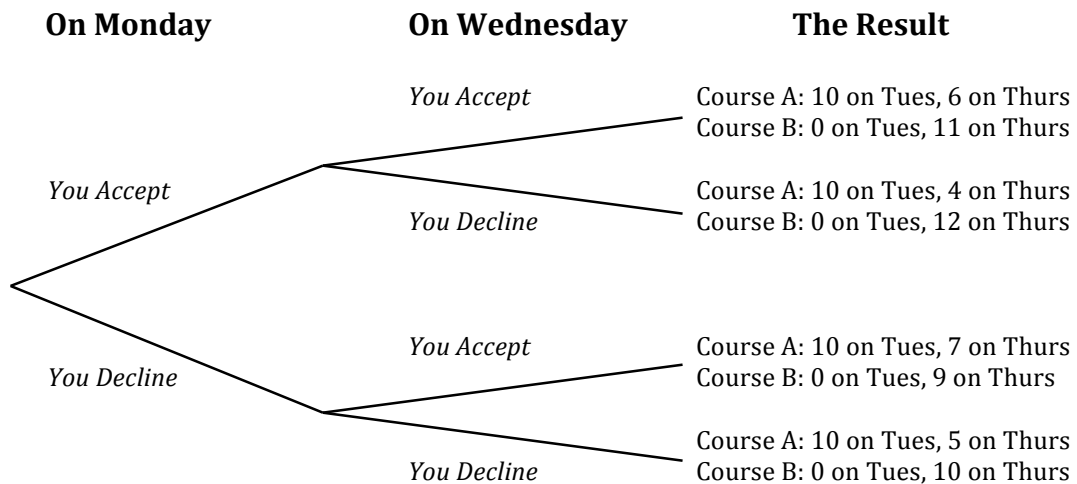
*really bad* things happening or having happened. But what things you consider *really bad* may depend on whether you are future-biased.

Dougherty's second observation is that sometimes a future-biased and risk averse person may be driven to act to her own acknowledged disadvantage. He imagines a situation of this general kind (I have changed some details):

#### Two Courses of Treatment

At all times you know all this: You are to scheduled to undergo one of two courses of treatment. Course A involves a painful, ten minute operation on Tuesday and another painful, five minute operation on Thursday. Course B involves just one painful, ten minute operation on Thursday. Either way, on Monday and Wednesday you will have no idea which course you are on (Wednesday-amnesia is part of both courses). And, either way, on Monday you will be offered a deal that will take a minute off the Thursday operation if you are on Course A, and add two minutes to the Thursday operation if you are on course B. And, either way, on Wednesday you will be offered a deal that will add two minutes to the Thursday operation if you are on course A, and take a minute off the Thursday operation if you are on course B. Your decision on Monday will have no causal influence over your decision on Wednesday.

What will you do, in this situation, if you are future-biased and risk averse? Note that there are four paths for you to go down over the course of the week, and the results of your going down these paths look like this:



On Monday, being risk-averse, you want to protect yourself against the thing you now consider to be *really bad* happening – your being on Course B. So you prefer *Accepting-then-Accepting* to *Declining-then-Accepting*, and you prefer *Accepting-then-Declining* to *Declining-then-Declining*. (In something closer to English: on Monday you would rather that you accept Monday’s deal, irrespective of what you will later do.) On Wednesday, being risk-averse and future-biased, you want to protect yourself against the thing you now consider to be really bad happening – your being on Course A. So you prefer *Accepting-then-Accepting* to *Accepting-then-Declining*, and you prefer *Declining-then-Accepting* to *Declining-then-Declining*. (On Wednesday you would rather that you accept Wednesday’s deal, irrespective of what you did earlier). But throughout you prefer *Declining-then-Declining* to *Accepting-then-Accepting*. (Throughout you would rather that you decline both deals than accept both deals).

In this situation, if you act on your preferences on Monday, and again act on your preferences on Wednesday, then you will be working to your own

acknowledged disadvantage – by condemning yourself to an extra minute of pain on Thursday, no matter which course you are on.

What should we make of this? The weak conclusion to draw is just that sometimes, when you lack the power to self-bind (in this case: the power on Monday to prevent your later self from accepting Wednesday's deal) then it is undesirable to be risk-averse and future-biased. This is no great news. For any attitude you might have, we can imagine situations in which it is undesirable to have that attitude. When your head will explode if you love your mother it is undesirable for you to love your mother. It would be rash to conclude that in normal, non-explosive contexts, there is something awry with you loving your mother.

Dougherty tentatively pushes us to a much stronger conclusion, the conclusion that it is a rational defect in you to be future-biased. The difference between his example and the loving-your-mother example is that in the former the outcome that is undesirable by your own lights comes about as a result of your own free choices, choices that you endorse throughout. You are acting in a disunified way. But rational people never act in disunified ways.

Why do rational people never act in disunified ways? Dougherty does not spell out the answer in detail. His basic reasoning, I take it, is this: in Two Courses of Treatment, if you are future-biased and risk-averse then, whatever you do, you are an appropriate subject of rational criticism. If, on Monday, you decline the first deal, then the critic can say “Why didn't you accept it? On Monday you preferred (irrespective of what you would do later) that you accept the first deal.” If, on Wednesday, you decline the second deal, then the critic can say “Why didn't you

accept it? On Wednesday you preferred (irrespective of what you did earlier) that you accept the second deal.” If you accept both deals then the critic can say “Why didn’t you decline both deals? At all times you preferred that you decline both deals than that you accept both deals.” But rational people are not appropriate subjects of rational criticism so, faced with this decision problem, if you are future-biased and risk-averse then you are irrational. But rational people are not such that they would be irrational if faced with the wrong decision problem. So, simpliciter, if you are future-biased and risk-averse then you are irrational.

Is the reasoning good? Here’s a piece of circumstantial evidence against it: There are other cases in which people, by acting on their preferences (endorsed throughout) in a step-by-step way, work to their own acknowledged disadvantage in much the same way as the future-biased, risk-averse person does in Two Courses of Treatment, and in some of these cases we do not want to say that the people are irrational in virtue of having bad preferences. Consider Satan’s Apple (from Arntzenius, Elga and Hawthorne 2004 – this is the diachronic version of their case):

#### Satan’s Apple

Satan cuts an apple into infinitely many slices and offers them to Eve, one by one, over the course of an hour – one at 11am, another at 11.30, another at 11.45... etc. Eve will make infinitely many decisions, knowing that no decision she makes will influence any later decision she makes. If, at noon, she has eaten infinitely many slices of apple, then she will Fall. Otherwise she will remain in Eden.

Eve strongly prefers Eden to Earth. And, all-other-things-being-equal-Eden-and-Earthwise, she prefers to eat more apple rather than less apple. What should she do?

It appears as if, whatever Eve does, she will be an appropriate object of rational criticism. If she fails to eat any one slice then the critic can say “Why didn’t you eat that slice? You preferred (irrespective of what you would later do) that you eat it, and you knew that your eating it would have no bearing on whether you ate finitely many or infinitely many slices.” If she eats every slice then the critic can say “Why did you eat infinitely many slices? You preferred that you eat finitely many slices.”

Whatever we conclude from this example, we do not want to conclude that there is something awry with Eve’s preferences – that we would all be rationally defective in preferring Eden to Earth, more apple to less apple.

Where exactly does the argument to this conclusion go wrong? This is a hard question. Let me put an answer on the table. If Eve takes all the slices then she is not an appropriate subject of rational criticism. Rational criticism is, directly or indirectly, criticism of *reasoning*. When I rationally criticize you for acting, or failing to act, a certain way, I indirectly criticize you for the reasoning that led you to decide, or fail to decide, to act that way.<sup>v</sup> And the criticism of your action is only as good as the criticism of your reasoning. So there is a constraint on appropriate rational criticism:

*Decision-Governed Accountability*

You are appropriately rationally criticized for acting, or failing to act, a certain way only if there is some time at which, if you were to decide at that time to act that way, then you would act that way.



But there is no time such that if Eve had decided at that time to take finitely many slices then she would have taken finitely many slices. By hypothesis her later decisions were causally isolated from her earlier decisions.<sup>vi</sup>

And the same can be said for you in the Two Courses of Treatment case. If you accept both deals then we cannot rationally criticize you for failing to decline both deals, because there was never a time such that, if you had decided at that time to decline both deals then you would have declined both deals. By hypothesis your Wednesday decision was causally isolated from your Monday decision.

## **6. So Where Does this Leave Us?**

In brief: We are all future-biased with respect to our own bad experiences, and this phenomenology of this preference is quite unlike the phenomenology of preferences that we take to be rationally optional – like a preference for strawberry ice cream over vanilla ice cream. To put it in a more precise way than the inchoate phenomenology may warrant, it seems to us that it *makes sense* to be future-biased. But, in spite of the efforts of numerous philosophers, it has proven to be extremely difficult to isolate good, persuasive arguments that support or undermine this seeming. This suggests that, if there are normative facts concerning future-bias they are *brutish* facts – facts that philosophy is ill-suited to uncover.

## References

- Arntzenius, Frank, Adam Elga and John Hawthorne. 2004. "Bayesianism, Infinite Decisions and Binding." *Mind* 113 (450): 251-283.
- Caruso, Eugene, Daniel Gilbert and Timothy Wilson. 2008. "A Wrinkle in Time: Asymmetric Valuation of Past and Future Events", *Psychological Science*, 19 No. 8: 796-801.
- Dougherty, Tom. 2011. "On Whether to Prefer Pain to Pass." *Ethics* 121, no. 3: 521-537.
- Hare, Caspar. 2008. "A Puzzle About Other-Directed Time-Bias." *Australasian Journal of Philosophy* 86, no. 2: 269-277.
- Hare, Caspar. 2009. *On Myself and Other, Less Important Subjects*. Princeton NJ: Princeton University Press.
- Hedden, Brian. Forthcoming. "Options and the Subjective Ought." *Philosophical Studies*.
- Suhler, Christopher and Craig Callender. "Thank Goodness that Argument is Over: Explaining the Temporal Value Asymmetry." *Philosopher's Imprint* forthcoming
- Maclaurin, James, and Heather Dyke. 2002. "Thank Goodness That's Over: The Evolutionary Explanation." *Ratio* 15: 276-292.
- Lewis, David. 1979. "Attitudes De Dicto and De Se". *The Philosophical Review* 88, no. 4: 513-543,
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press
- Prior, A.N. 1959. "Thank Goodness That's Over." *Philosophy* 34 No. 128: 12-17

---

<sup>i</sup> Psychologists and economists have paid a great deal of attention to what philosophers call 'near-bias' – our tendency to care less about far-future goods and bads than about near-future goods and bads. But that is a different thing.

<sup>ii</sup> The basic psychological mechanism behind future-bias has recently received some attention -- see Caruso, Gilbert and Wilson (2008), and Suhler and Callender forthcoming. But those authors do not focus on what kinds of disvalued events are objects of future-bias.

<sup>iii</sup> The growing block theory of time (which says that past events exist, future events do not) has had some supporters. See C.D. Broad *Scientific Thought*, Harcourt Brace and Co. 1923, and Michael Tooley *Time, Tense and Causation*, Oxford University Press 1997. But there is at least something that might be said for it – it somehow bears out our inchoate idea that the future is open and the past closed.

<sup>iv</sup> A similar sort of worry arises for distant-stranger benevolence on the assumption that Lewisian modal realism is true. If modal realism is true then my desire that distant strangers not starve amounts to a preference that I be non-worldmate related to starving strangers. Buy why care about that? It is not okay just to say "it matters, I can't explain why."

<sup>v</sup> A stronger view, according to which the proper objects of rational assessment are *decisions*, not *actions*, has recently been robustly defended by Brian Hedden – (see Hedden forthcoming).

<sup>vi</sup> Of course, as Tom Dougherty has pointed out, this entails that predictable exploitability is not always a mark of irrationality. So, if we take this line, we cannot say, for example, that the predictable exploitability of people with intransitive preferences shows them to be irrational. That's a cost we must bear.