Why are people so darn past biased?

Abstract

Many philosophers have assumed that our preferences regarding hedonic events exhibit a bias toward the future: we prefer positive experiences to be in our future and negative experiences to be in our past. Recent experimental work by Greene et al. (forthcoming) confirmed this assumption. However, they noted a potential for some participants to respond in a deviant manner, and hence for their methodology to underestimate the percentage of people who are time neutral, and overestimate the percentage who are future biased. We aimed to replicate their study using an alternative methodology that ensures there are no such deviant responses, and hence more accurately tracks future bias and time neutrality. Instead of finding more time neutrality than Greene et al., however, we found vastly more past bias. Our explanation for this surprising finding helps to reveal the rationale behind both future and past biased preferences, and undermines the generalisability of one of the most influential motivations for the rationality of hedonic future bias: Parfit’s My Past or Future Operations.

1. Introduction

We will say that an agent is biased toward the future if they tend to prefer that positively valenced events be in their future and that negatively valenced events be in their past. When these preferences concern sensations, such as pleasure or pain, an agent is said to be hedonically future biased.¹

There has been much debate between philosophers regarding the rationality of hedonic future bias. It has been argued that such preferences are rationally required, that they are

¹ Greene and Sullivan (2015, 948–9) provide a more formal characterisation of hedonic time biases.
rationally permissible but not required, and that they are irrational. This normative disagreement is accompanied by widespread agreement regarding the preferences people in fact have. It is agreed that people are in fact hedonically future biased; what is in dispute is the extent to which these preferences are rational.

Following recent work by Greene, Latham, Miller and Norton (forthcoming), we will distinguish between positive hedonic future bias—the preference to have pleasures in one’s future rather than one’s past—and negative hedonic future bias—the preference to have pains in one’s past rather than one’s future. Notably, those who argue that hedonic future bias is rational typically focus on the latter kind of case (see for instance Prior (1959), Parfit (1984, 165), Hare (2007; 2008), and Heathwood (2008)).

Future bias is contrasted with time neutrality and past bias. People are time neutral concerning some type of event if they have no preference whether it is in their future or past. People who are hedonically past biased tend to prefer that positively valenced hedonic events be in their past and that negatively valenced hedonic events be in their future. No philosophers posit that people are hedonically past biased.

Greene et al. aimed to test philosophers’ predictions about people’s preferences concerning various kinds of events. As predicted, they found that participants

---


3 While philosophers have assumed that people are hedonically future biased, they have supposed that these preferences are accompanied by time neutral preferences regarding both non-hedonic events (events that are not directly experienced, or where the experience is not linked to the temporal location of the event) and third-person events (hedonic or non-hedonic events that are experienced by, or relevant to, someone else). In other words, it is assumed that if the event in question is non-hedonic, or concerns some third party, people will have no preferences regarding whether it is in the future or past. For discussion see Brink (2011), Hare (2013), Dougherty (2015), and Greene and Sullivan (2015).

4 Indeed, the term ‘past bias’ is not one that appears in the literature on time bias. However, the concept is clear—it is simply the inverse of future bias—and given the results presented in this chapter, it is a concept in need of consideration.
demonstrated both negative and positive hedonic future bias. They noted, however, one limitation of their study. Participants read a vignette and were then presented with a statement of the form ‘I would prefer to learn that [the event] occurred yesterday’, or ‘I would prefer to learn that [the event] will occur tomorrow.’ They were then asked to respond to that statement on a Likert scale ranging from 1 ‘strongly disagree’ to 7 ‘strongly agree’.

The starting assumption of the experiment was that participants would choose 4 to reflect indifference, and 1 or 7 to indicate either future bias or past bias. However, Greene et al. noted that they could not rule out that a small percentage of strongly time-neutral participants might respond in a deviant manner, by using 1—strongly disagree—to express their time neutrality (by strongly disagreeing that they have a preference that the event be located in their past/future). They point out that if there are time neutral people responding in this manner, their methodology might lead to an overestimation of participants’ past bias and future bias, and an underestimation of their time neutrality.

Our aim in the current study was to see whether this was so, or whether people really are as hedonically future biased as the Greene et al. study found them to be. To that end we used the same vignettes that they used in their study, but our methodology allowed no room for deviant responses. Rather than being asked to agree or disagree with a statement, participants were able to directly indicate their preference by moving a slider on a Likert scale between ‘I would strongly prefer to learn that [the event] occurred yesterday’ and ‘I would strongly prefer to learn that [the event] will occur tomorrow’ via a mid-point of ‘I have no preference between these two options.

Greene et al.’s other results are more surprising. They found that there are no general hedonic/non-hedonic and first/third-person asymmetries in temporal preferences, which undermines arguments for the irrationality of future bias that appeal to such asymmetries. Nevertheless, they did observe that non-hedonic and third-person conditions tend to decrease future bias. They suggested that there are at least three dimensions that influence whether preferences are future biased or temporally neutral. Future bias tends to increase when events are negatively valenced, hedonic, or first-person, and it tends to decrease when events are positively valenced, non-hedonic, or third-person.
Since this study involved the same vignettes and only a small change to the way in which participants were asked to indicate their preference, we predicted that the current study would replicate the same general pattern of results as that found by Greene et al.; that is, we predicted that people would be both negatively and positively hedonically future biased. This is not what we found. While we replicated Greene et al.’s finding that people prefer positively valenced events in their future rather than their past (positive hedonic future bias) we did not replicate their finding that people prefer negatively valenced events in their past rather than their future (negative hedonic future bias). Instead, we found that people preferred to have negatively valenced hedonic events in their future rather than their past: they exhibit negative hedonic past bias. Thus, we found that people preferred to have both positive and negative events in their future rather than their past.

In §3 we propose an explanation of our findings and consider the upshots for arguments regarding the descriptive realities and normative status of future bias. By comparing our results to those of Greene et al., we are able to rule out some prominent and otherwise tempting hypotheses for why people have future biased and past biased preferences: the bad memories and improving sequences hypotheses. The comparison also undermines the generalisability of one of the most celebrated and influential motivations for negative hedonic future bias: Parfit’s My Past or Future Operations thought experiment.

2. Experimental Design and Results

2.1 Method

2.1.1 Participants

153 people participated in the study. Participants were U.S. residents, recruited and tested online using Amazon Mechanical Turk, and compensated $0.50 for approximately 5 minutes of their time. 52 participants had to be excluded for failing to follow task
instructions. This means that they failed to answer the questions (36), or failed an attentional check question (16). The remaining sample was composed of 101 participants (aged 23-66; 39 female; mean age 30.59 (SD = 8.79)). Ethics approval for this study was obtained from the University of Sydney Human Research Ethics Committee. Informed consent was obtained from all participants prior to testing. The survey was conducted online using Qualtrics.

2.1.2 Materials and Procedure

Participants were randomly assigned to one of two conditions: positive hedonic or negative hedonic. We hypothesised that we would find future bias in both conditions.

We used Greene et al.’s first-person positive and negative hedonic vignettes. Each involves receiving a meal, and the only difference between them concerns whether the experience is a positive hedonic experience (receiving one’s favourite meal) or negative hedonic experience (receiving one’s most disliked meal).

Thus, participants read a version of the following vignette:

Imagine you are an astronaut on a 10-year voyage between planets. You are 5 years into the voyage. The ship’s food dispenser normally produces bland meals containing only essential nutrients. However, it is programmed to dispense your [favourite]/[most disliked] meal — which you really [like]/[dislike] — during one day of the voyage. One morning, you awake from a dream concerning your [favourite]/[most disliked] meal and for a moment you cannot remember whether you have received it yet.

In each condition, participants were presented with a Likert scale that ran from 1 ‘I would strongly prefer to learn that my [favourite]/[most disliked] meal was dispensed yesterday, and will not be dispensed tomorrow’ at one end (either the far left or the far
right, determined randomly) to 7 ‘I would strongly prefer to learn that my
[favourite]/[most disliked] meal will be dispensed tomorrow, and was not dispensed
eyesterday’) at the opposite end of the scale via 4 ‘I have no preference between these two
options’, in the middle of the scale. Participants then slid the slider to wherever on the
scale they took to reflect their preference.

In both conditions participants were then asked to indicate their level of confidence in
their previous judgement.6 After having done so, participants answered a comprehension
question: “In this vignette, you were asked to imagine that you were…” to which they could
answer (1) an astronaut; (2) a dog; (3) a builder or (4) the home-town mayor. Participants
who did not choose (1) were excluded.

2.1.3 Analyses

In order to compare the results between conditions, we re-coded participants’ responses
in such a way that a response of 5, 6, or 7 reflects future bias (a preference that the event
in question be future if it is positive and past if it is negative) in both conditions. A
response of 1, 2, or 3 reflects past bias (a preference that the event in question be future
if it is negative and past if it is positive) in both conditions. A response of 4 indicates
time neutrality.

In order to test for future bias we first ran separate one-sample t-tests to test whether the
mean response significantly differs from 4 in each condition. If the mean is significantly
above 4, then overall people might be future biased; if the mean is significantly below 4
then overall people might be past biased; if the mean does not differ significantly from 4
then overall people might be time neutral. (We say might be, here, since it is consistent
with there being a mean significantly above 4, that we find a population of future biased

---

6 Participants were quite confident in their judgements regarding their preferences ($M = 5.48, SD = 1.15$). Results of a between-subjects t-test showed that there was no significant difference in level of confidence between conditions ($p = .207$).
participants, and a slightly smaller population of past biased participants; this distribution would not vindicate people overall being future biased. *Mutatis mutandis* for a mean that does not significantly differ from 4, which might reflect a large population of time neutral participants, or two equal populations of future biased and past biased participants). We compared future bias between the two conditions with a between-subjects t-test.

In the condition in which the mean is significantly greater than 4, we combined the proportion of people who were past biased with those who were time neutral—we will call these people non-future-biased. In the condition in which the mean is significantly less than 4, we combined the proportion of people who were future biased with those who were time neutral—we call these people non-past-biased. We then ran separate one-way $\chi^2$-tests to test whether in the former case, the majority of people responded in a future biased manner, and in the latter case, the majority of people responded in a past biased manner.

### 2.2 Results

We hypothesised that (i) we would find future bias in the positive hedonic condition and (ii) we would find future bias in the negative hedonic condition. The first hypothesis was confirmed: we found a majority of participants to be future biased in the positive hedonic condition. However, the second hypothesis was quite spectacularly disconfirmed: surprisingly, we found a majority of participants to be past biased in the negative hedonic condition.

Table 1 below summarises the descriptive data from the experiment. After the re-coding described in §2.1.3, the ‘FB’ column represents the proportion of participants who reported future biased preferences (5, 6 or 7) and the ‘PB’ column represents the
proportion of participants who reported past biased preferences (1, 2 or 3). The ‘N’ column represents the proportion of people who reported time neutrality (4).

Table 1. Descriptive data and one-sample t-test results from both conditions.

<table>
<thead>
<tr>
<th>Condition</th>
<th>%FB</th>
<th>%PB</th>
<th>%N</th>
<th>Mean</th>
<th>SD</th>
<th>t-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1: Negative Hedonic (N = 49)</td>
<td>18.4</td>
<td>69.4</td>
<td>12.2</td>
<td>3.14</td>
<td>1.28</td>
<td>-4.707</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Condition 2: Positive Hedonic (N = 52)</td>
<td>67.3</td>
<td>15.4</td>
<td>17.3</td>
<td>5.10</td>
<td>1.45</td>
<td>5.469</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

The results of our one-sample t-tests (which, recall, show us whether the mean response differs significantly from a value of 4) appear to show that overall, people are future biased in the positive condition and past biased in the negative condition. Unsurprisingly, then, the results of a between-subjects t-test confirm that people are significantly more future biased in the positive condition than in the negative condition ($t(99) = -7.186, p < .001$).

However, the one-sample t-test does not tell us whether the majority of people in a given condition are in fact past biased, or future biased: for that we must look to the results of our one-way $\chi^2$-tests, reported in Table 2 below. As per §2.1.3, in the negative condition, where we found overall past bias, we grouped together time neutral and future biased participants as a non-past-biased group (non-PB). In the positive condition, where we found overall future bias, we grouped together time neutral and past biased participants as a non-future-biased group (non-FB).

Table 2. Results of one-way $\chi^2$-tests.

<table>
<thead>
<tr>
<th>Condition</th>
<th>%PB</th>
<th>%non-PB</th>
<th>$\chi^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Condition 1: Negative Hedonic</td>
<td>69.4</td>
<td>30.6</td>
<td>7.367</td>
<td>.007</td>
</tr>
<tr>
<td>Condition 2: Positive Hedonic</td>
<td>67.3</td>
<td>32.7</td>
<td>6.231</td>
<td>.013</td>
</tr>
</tbody>
</table>
Recall that our one-way $\chi^2$-tests show us whether the proportion of participants responding in a past biased versus non-past-biased way (in the negative hedonic condition) or future biased versus non-future-biased way (in the positive hedonic condition) differs significantly from a 50/50 split. Our results show that a significant majority of people were past biased in the negative hedonic condition and the majority of people were future biased in the positive hedonic condition.

3. Discussion

Based on prior studies and the assumptions of philosophers, we predicted that participants would exhibit future bias in both the negative and positive hedonic conditions. While we indeed found future bias in the positive hedonic condition, the negative hedonic condition revealed a strong past bias.

Finding past biased responses to negative hedonic events is surprising, in itself, but it is even more so given that it resulted from a small change in how participants are asked to indicate their preference. Recall that in Greene et al.’s prior study, participants were prompted to agree or disagree with a statement of the form ‘I would prefer to learn that my most disliked meal will be dispensed [tomorrow/yesterday], and was not dispensed [yesterday/tomorrow]’ on a Likert scale from 1 ‘strongly disagree’ to 7 ‘strongly agree’. In the current experiment, participants were also presented with a 7-point Likert scale, ranging from ‘I would strongly prefer to learn that my most disliked meal was dispensed yesterday, and will not be dispensed tomorrow’ at one end (either the far left or the far right, determined randomly) and ‘I would strongly prefer to learn that my most disliked
meal will be dispensed tomorrow, and was not dispensed yesterday’ on the other end, via ‘I have no preference between these two options’ in the middle.

The only difference between the two methodologies is the way in which participants indicate their preference. No changes were made to the vignettes themselves. Indeed, as expected, our modification to the way in which participants indicate their preferences produced little difference in the positive condition, in which a strong future bias was observed in both studies.

The fact that participants respond in a strongly past biased way when asked to indicate their preference in one way, and a strongly future biased way when asked to indicate in another way, allows us the opportunity to study the mechanisms, or rationale, behind future biased and past biased preferences.

3.1 Pure Time Preference

When philosophers debate the rationality of hedonic future bias they typically have in mind “pure” hedonically future biased preferences. As Lowry and Peterson (2011, 490) write, “Pure time preference is a preference for something to come at one point in time rather than another, not because this will make the benefit greater or more certain, but merely because of when it occurs in time.” Pure hedonically future biased preferences are a response to the temporal location of an experience and not to the fact that one option probably entails more positive experiences or less negative experiences overall (from a time neutral perspective). For example, someone with pure negative hedonically future biased preferences prefers negative experiences to be past, not because this will make the experiences less bad from a time neutral perspective, but merely because of where the negative experience occurs in time.
If a person prefers that an event occur in the future or the past, and their reasoning is based entirely on the fact that the total utility they receive will be greatest if the event occurs in that location (from a time neutral perspective), then that person can be said to be time neutral about the event in question. They care about which option will entail the most good and least bad for their life overall, and not which option puts good things in their future and bad things in their past. Thus, such a person would not display a pure future or past biased preference.

There are at least two more factors, other than total utility, that philosophers claim affect time preferences but do not qualify as pure time preference: considerations of personal identity and improving sequences. For example, in attacking the rationality of pure hedonic time bias, Greene and Sullivan (2015, 948, fn. 1) write,

> First, an agent might prefer near pleasures and distant pains for reasons concerning personal identity (e.g., the agent might be skeptical that she will be the same person that feels distant pain and pleasure). Second, an agent might prefer that pleasures and pains occupy certain temporal locations because she has “global” preferences about the structure of her life (e.g., she prefers an improving life to a degrading one, regardless of the total amount of pleasure in each life).

We remain neutral regarding the rationality of these types of preferences.

As with considerations of total utility, if a person displays a preference for an event to be future or past, and in doing so they are reacting only to considerations of personal identity or improving sequences, then that person does not display a pure time biased preference. Their motivation is consistent with time neutrality.
3.2 The Mitigation Hypothesis

Our hypothesis is that participants in our negative hedonic condition do not display pure past biased preferences; instead, we suspect that they are reacting to considerations of total utility in a time neutral way. This is a fascinating hypothesis because it entails that we were able to induce time neutrality in participants by making only a small methodological change to the way in which they report their preference. Therefore, if true, this explanation would give us a valuable insight into the mechanisms behind pure time preference and time neutrality.

Before turning to our preferred explanation, let us first consider two other tempting hypotheses in the philosophical literature that might be used to claim that seemingly future biased or past biased participants are actually responding in a time neutral way. By comparing Greene et al.’s earlier results with ours, we are able to rule out these hypotheses.

First, it might be hypothesised that considerations of total utility explain past biased responses because negative events that occur in the past can affect the agent in a negative way for a longer amount of time. This factor is particularly salient when we consider events that are not simple positive or negative hedonic experiences. For example, some people might hold that the total utility of their lives is increased if events that negatively influence reputation are in their future. If the consequences of such an event persist throughout one’s life, then some people may reason that by locating these events in the future, less of their total life is negatively affected by the reputational cost than if that event were located in the past. Thus, if they want to minimise the total utility lost as a result of these reputational consequences, they will prefer to have these events located in the future, and hence will be past biased about these kinds of events.
Similarly, it might be hypothesised that participants want their negative hedonic experiences to be in the future because that entails that the total amount of one’s life that is marred by the negative influence of this event is lessened. Specifically, there is less time for the memory of the least favourite meal, or other effects (on the health of one’s gut, for example), to have an influence. Call this the bad memories hypothesis.

A comparison to Greene et al.’s earlier study rules out the bad memories hypothesis as a satisfactory explanation of negative hedonic past biased preferences. In that study, responses were consistent with a strong negative hedonic future bias, and the vignettes were identical. If people are being motivated by a preference to minimise the total amount of time that a negative experience has an effect on them, then we would expect this to be the case in both experiments, because in each case the bad memories hypothesis has equal applicability. Therefore, the bad memories hypothesis cannot explain why the way in which participants are asked to indicate their preference would have a large effect on their preferences.

A related hypothesis that might be used to explain people’s preferences is the improving sequences hypothesis. It has long been hypothesised that individuals care not just about the total utility of their lives, but about the distribution of that utility. Indeed, some philosophers, such as Velleman (1991) and Slote (1983, 23-24) hold not only that people have such preferences, that but they are rational to do so. Philosophers like Slote and Velleman claim that people care about the distribution of utility rather than its aggregate, and they favour improving sequences. For example, Slote and Velleman claim that an improving life is better than a worsening life even if the ‘area under the curve’ is the same; i.e., even if the aggregate utility is the same, (or, perhaps even if it is marginally worse).
Such a consideration would pull in favour of future biased responses for both negative and positive conditions, as was observed in Greene et al.’s earlier study. However, a comparison to our study rules out the *improving sequences hypothesis* as a satisfactory explanation for positive hedonic future biased responses. In our study, responses were strongly past biased in the negative hedonic condition, and the vignettes were identical with those used by Greene et al. If people favour improving sequences, then we would expect this to be the case in both experiments, because this consideration is equally applicable in each experiment. Therefore, the improving sequences hypothesis cannot explain why the way in which participants are asked to indicate their preference would have a large effect on their preferences.

Thus we think that neither the bad memories hypothesis nor the improving sequences hypothesis can adequately explain the results of both experiments. We think that a more satisfactory explanation for the observed results is one that appeals to participants’ conceptions of their own agency and ability to intervene on future events. People have a general sense that the future is open, and the past closed or fixed. What is done is done, but what is future is still open to intervention (see Ismael 2012 for discussion of such a view). Some ways of asking participants to report their preferences encourage participants to take an active role in considering ways they might intervene on events and other ways tend to discourage it. Asking participants to move a slider to indicate their preference for an event to be past or future might encourage participants to imagine that they are *placing* the event in the past or future; that they have a *choice* in the temporal location of the event.

For example, it might be that participants are reasoning that even if they cannot intervene on the schedule by which the meals are delivered by the computer, if they choose to place their most disliked meal tomorrow, they can also choose to save and set
aside some of today’s bland meal to eat tomorrow, in order to mitigate the badness of receiving the disliked meal. Perhaps, in addition, they assume that if they place the disliked meal in the past, they would not have had an opportunity to intervene in this way (because they did not know that the most disliked meal was coming).

It might be, then, that in conditions in which people think that it is within their control to mitigate bad events, some people will prefer to have the negative event in the future in order to facilitate the opportunity to engage in that mitigation. Then these people might respond in an apparently past biased manner, but because they take themselves to be reporting a preference for less future pain over more past pain, this need not indicate pure past biased preferences. Call this the mitigation hypothesis.

Notably, that explanation is consistent with there being some people who we report as past biased, who nonetheless prefer, all else being equal, to have negative hedonic events located in their past (people who have purely negatively hedonically future biased preferences) due to the sorts of considerations already adumbrated by philosophers. We can expect this whenever participants are encouraged to view trade-offs in a fixed-future/fixed-past manner. If participants are encouraged to view future experiences as just as fixed as past ones—that is, the experiences associated with the events will be exactly as bad whether they occur in the past or future—then we would expect such participants to display pure hedonically future biased preferences in the negative hedonic condition.

Asking participants to agree or disagree with statements about the preferability of the most disliked meal occurring in the future/past, as per Greene et al.’s methodology, encourages participants to view the situation in a fixed-future/fixed-past way: participants are less likely to engage with the possibility of mitigating these events because they are not taking an active role in “placing” the events temporally.
In general, then, when participants view a trade-off in an open-future/fixed-past way, we expect to find that some participants who have an attitude of time neutrality, or future bias, to negative hedonic events, will nonetheless respond in a past biased manner when reporting preferences about a bad event whose badness can be mitigated only if it is in the future. When participants view a trade-off in a fixed-future/fixed-past way (or the events in question are difficult or impossible to mitigate), we expect to find larger populations of future biased responses in negative conditions.7

3.3 Implications for the Debate over the Rationality of Hedonic Future Bias

Does the mitigation hypothesis bear on the debate over the normative status of hedonic future bias? Consider first the distinction between absolute and non-absolute hedonic future bias.8 An absolutely negatively future biased agent doesn’t care about past negative experiences at all. Such an agent will always prefer a negative experience to be past rather than future, even if the past experience is many times worse than the future experience. Non-absolutely negatively future biased agents will sometimes prefer future negative experiences to past ones, if the differential between the two is great enough.

The results of the current experiment do seem to rule out that people are, in general, absolutely negatively future biased. Even if people reason that they can mitigate the badness of the experiences associated with receiving one’s most disliked meal in the future, they still must deal with some negative future experiences—at a minimum, they must either eat some of their most disliked meal, go hungry tomorrow, or go somewhat hungry both today and tomorrow by saving some of today’s bland meal to eat tomorrow. In contrast, if the event is in the past, then whatever negative experiences they had to undergo are over and done with. Therefore, if people were absolutely negatively future

---

7 By similar reasoning, in the positive condition participants would presumably want to enhance positive experiences, in which case the open-future/fixed-past idea would encourage future biased responses in positive conditions, which is what we observed.

biased, then we would expect them to prefer that the event be in the past, regardless of whether or not they could mitigate its badness by placing it in the future.

However, our results do not rule out the possibility that people are non-absolutely negatively future biased. Perhaps people are negatively hedonically future biased, but this future bias can be overridden by considerations of total utility when the differential between negative future and past experiences is great enough. This leaves open two possibilities: i) that participants in our experiment continued to possess a non-absolute hedonic negative future bias, but this was overcome by considerations of total utility when the possibility of mitigation was made salient, and ii) the possibility of mitigation caused participants to change from a future biased perspective to a time neutral one.

Even though our results do not rule out (i), on reflection, we believe that (ii) is more plausible. We suggest that participants change from a negatively hedonically future biased perspective to a time neutral one when it is made salient to them that they can intervene on negative events and mitigate the badness of the experiences associated with them. That is, participants are much more likely to adopt a time neutral perspective when they view future negative experiences as being less bad because of their own agency.

We can test this explanation by considering a choice between a less bad negative future experience and a more bad negative past experience, and comparing cases in which the agent has no control over the situation against cases in which the future experience is less bad because the agent has taken an active role in making it less bad.

Luckily, such an opportunity presents itself in what is arguably the most influential thought experiment in the history of the debate over hedonic future bias: Parfit’s *My Past or Future Operations*. Parfit writes:
I am in some hospital, to have some kind of surgery. Since this is completely safe, and always successful, I have no fears about the effects. The surgery may be brief, or it may instead take a long time. Because I have to co-operate with the surgeon, I cannot have anaesthetics. I have had this surgery once before, and I can remember how painful it is. Under the new policy, because the operation is so painful, patients are now afterwards made to forget it. Some drug removes their memories of the last few hours.

I have just woken up. I cannot remember going to sleep. I ask my nurse if it has been decided when my operation is to be, and how long it must take. She says that she knows the facts about both me and another patient, but that she cannot remember which facts apply to whom. She can tell me only that the following is true. I may be the patient who had his operation yesterday. In that case, my operation was the longest ever performed, lasting ten hours. I may instead be the patient who is to have a short operation later today. It is either true that I did suffer for ten hours, or true that I shall suffer for one hour.

I ask the nurse to find out which is true. While she is away, it is clear to me which I prefer to be true. If I learn that the first is true, I shall be greatly relieved.

With this thought experiment, Parfit hopes to elicit the intuition that it is preferable to have a much more painful surgery in one’s past than to have a much less painful surgery in one’s future. If this intuition is widespread, it would support either absolute negative hedonic future bias, or at least a particularly strong form of non-absolute negative hedonic future bias. And, indeed, it seems that Parfit succeeds in showcasing this intuition.
However, Parfit’s case is also particularly adept at suppressing the agency of the protagonist. In the thought experiment, Parfit is lying in a hospital bed, with amnesia, and his understanding of the situation is completely dependent on the information to be provided to him by the nurse. There is no reason to suppose that he can take an active role in mitigating anything that might happen to him; he will either be in pain in the future or was in pain in the past, depending on what he learns is true. Parfit is simply reporting his preference for which statement about his situation he would most like to hear from the nurse. This is the way that preferences were collected in Greene et al.’s (forthcoming) prior experiment that showed negative hedonic future bias.

Imagine, now, a case in which your agency, and ability to mitigate bad events, is restored. You just underwent an extremely painful surgery and are now lying in a hospital bed. You were told that you needed to take a pill approximately eight hours before the surgery to help mitigate the painful effects of the surgery, but this morning you forgot to do so. Before heading to the hospital, you called to ask whether the surgery could be delayed by a day, so you could take the pill. Unfortunately, the hospital refused, and so you were compelled to go ahead with the surgery, even though your failure to take the pill resulted in the surgery being ten times more painful than if you had remembered to take it.

Now that the surgery has concluded, what is your preference? Are you satisfied with your situation, because all your pain is now in the past? Or would you prefer that the hospital had allowed you to do the surgery tomorrow, so that you could take the anesthetic pill, thus reducing your pain by a factor of ten?

If there is an intuition in favor of time neutrality in this case, or even if the intuition in favor of future bias is simply less clear than it is in Parfit’s case, then that is strong evidence in support of the hypothesis that feelings of control shift our perspective from negative hedonic future bias to time neutrality. Notice that in each case—Parfit’s original
and our variant—the facts about whether pain will be past or future, and how much worse the pain will be, are identical. The only relevant difference is that in Parfit’s case, the agency of the protagonist in mitigating the badness of the future outcome is suppressed, while in our case, it is enhanced.

This explanation undermines the degree to which My Past or Future Operations can be generalised to other hedonic trade-offs. In doing so, it may be seen to support arguments against the rationality of hedonic future bias. Philosophers already note that future bias seems to be absent when the events in question are non-hedonic or when we form preferences about what happens to another person, and supporters of time neutrality have appealed to these asymmetries in claiming that hedonic future bias is arbitrary and therefore suspect as a rational response. We have suggested that hedonic future bias itself may be limited in scope. In particular, we have suggested that people tend to view a decision situation from a future biased perspective when they have no agency in mitigating the badness of the experiences in question, but they tend to adopt a time neutral perspective when such agency is emphasised. If so, then this may represent an additional reason to doubt that pure hedonic future bias is a rational response.

4. Conclusion

In all, our results suggest that time-biased preferences are much more complex than has been assumed. The philosophical literature assumes that people exhibit a hedonic negative future bias. We have suggested that when agents view a decision situation in an open-future/fixed-past way, they are likely to engage with the possibility of mitigating negative future experiences and adopt a time neutral perspective. From a time neutral perspective, it often makes sense to place negative events in one’s future, so that one has

---

9 See footnote 3 for philosophers’ predictions, and footnote 5 for Greene et al.’s (ms) empirical results.
the chance to mitigate the badness of the experiences associated with the event, and thus maximise one’s total utility. This results in participants reporting an impure past biased preference: participants want to place negative events in the future only because that will increase their total utility (from a time neutral perspective).

Thus, the mitigation hypothesis explains the mechanism behind negative hedonic past biased responses. It predicts that people will be more likely to adopt a time neutral perspective, and provide past biased responses, when they think that negative events are influenced by their own agency and are within their power to mitigate. In contrast, it predicts that people are more likely to adopt a future biased perspective when they think that negative events are not within their control and impossible to mitigate. It would be beneficial for future research to test this hypothesis in greater detail.

Parfit’s *My Past or Future Operations* provides a particularly apt test case for the hypothesis. We presented both Parfit’s original case and a variant in which the protagonist’s agency and ability to mitigate future negative events are emphasised. In each case, the difference in pain between events and their relative location in time are identical; the only relevant change is the agent’s influence over the situation. Therefore, if future bias is enhanced by Parfit’s formulation, and time neutrality by our variant, as we claim, then this would provide powerful evidence in favour of the mitigation hypothesis. It would also show that the evidence provided by Parfit’s thought experiment is narrower in scope than previously thought: *My Past or Future Operations* motivates negative hedonic future bias for only a certain type of hedonic trade-off, and does not generalise to others.

**References**


