DOI: 10.1111/phpr.12819

ORIGINAL ARTICLE

Philosophy and Phenomenological Research

On Preferring that Overall, Things are Worse: Future-Bias and Unequal Payoffs

Preston Greene[1] | **Andrew J. Latham**[2] | **Kristie Miller**[2] | **James Norton**[3]

[1]Nanyang Technological University, Singapore

[2]University of Sydney

[3]University of Iceland

Correspondence
Kristie Miller, University of Sydney.
kristie_miller@yahoo.com

Funding information
Australian Research Council, Grant/Award Number: DP180100105 and FT170100262

Abstract

Philosophers vigorously debate the rationality of *hedonic bias toward the future*: a systematic preference for pleasurable experiences to be future and painful experiences to be past. The debate over future bias is distinctive in philosophy because arguments made on both sides concern descriptive and empirically tractable claims about patterns of preferences and the psychological mechanisms that could explain these patterns. Most notably, philosophers predict that this bias is strong enough to apply to *unequal payoffs*: people often prefer less pleasurable future experiences to more pleasurable past ones, and more painful past experiences to less painful future ones. They also predict that future-bias is restricted to *first-person preferences,* and that people's *third-person preferences* are time-neutral. These claims feature in arguments both for and against the rational permissibility of future bias. Thus, we aimed to test whether these claims are descriptively accurate. Among our discoveries, we found that the predicted asymmetry between first- and third-person conditions is absent, and so cannot support arguments against the rationality of future-bias. We also uncovered an asymmetry between positive and negative events that might ground a new argument in favour of time-neutralism.

1 | INTRODUCTION

A person is *biased toward the future* when they tend to prefer positively valenced events to be in their future (positive future-bias) or negatively valenced events to be in their past (negative future-bias), all else equal. When the events in question concern sensations, such as pleasure or pain, an agent is *hedonically* future-biased.[1] In investigating hedonic future-bias, there are two important and distinct lines of inquiry: i) in what ways, and to what extent, are people future-biased? and ii) is future-bias rational?

Even before empirical investigation of the first line of inquiry, a normative disagreement had erupted between two groups of philosophers regarding the rationality of these preferences. Supporters of future-bias claim that it is rationally permissible, or even obligatory, while *time-neutralists* claim that future-bias is irrational.[2] Part of what makes this debate interesting is that the arguments made by both sides concern descriptive and empirically tractable claims about patterns of preferences and the psychological mechanisms that could explain these patterns. Supporters of future-bias, for example, typically claim that future-bias is the result of emotional responses that are widely shared, and that these emotional responses are rationally permissible.[3]

Time-neutralists, meanwhile, argue from the alleged fact that people are future-biased about some kinds of events but not others to the conclusion that hedonic future-bias is an evolutionary heuristic that is undermined by rational reflection. The idea, here, is that the kinds of preferences that are allegedly time-neutral (non-hedonic preferences and third-person preferences[4]) are precisely those where we ought expect more rationality, whereas those preferences thought to be informed by future-bias (i.e., first-person hedonic preferences) are those where we ought expect *bias* (in the pejorative sense).

The most likely cases in which future-bias will determine people's preferences are in what we call *conditions of equality* (or equal conditions). These are conditions in which people are asked whether they have a preference regarding having *N* units of pleasure (or pain) in the future, versus *N* units in the past. In these cases, the overall amount of pleasure, or pain, that the person experiences is the same either way: all that varies is where those sensations occur relative to the present (in the past versus in the future). The presence of future-bias in such conditions has been empirically verified by recent studies.[5]

While the question of whether future-bias is rational in conditions of equality is an interesting and important one, we think it more pressing to consider the status of future-bias in *conditions of inequality* (or unequal conditions). These are conditions in which people are asked whether they have

[1]For a more formal characterisation of hedonic future-bias see Greene and Sullivan (2015, 948–9).

[2]Explicit supporters of the rationality of hedonic future-bias include Prior (1959), Hare (2007; 2008), Heathwood (2008), and Dorsey (2018). Recent critics of hedonic future-bias include Brink (2011), Greene and Sullivan (2015), and Dougherty (2015). Hedden (2015) argues that hedonic future-bias is merely rationally permissible while Parfit (1984) remains neutral about its rationality. Kauppinen (2018) claims that hedonic future-bias is rationally permissible only when the agent never acts on their preference.

[3]See Greene and Sullivan (2015, 967–8) for discussion of this argument.

[4]Non-hedonic preferences concern events that make someone's life go better or worse, but involve no pain or pleasure, or where the temporal location of the pain or pleasure is divorced from the temporal location of the event itself. See Hare (2013) for examples and discussion. More on third-person preferences below.

[5]See Caruso, Gilbert, and Wilson (2008) and Greene, Latham, Miller and Norton (2021).

a preference regarding having *N* units of pleasure (or pain) in the future, versus *N** units in the past, where *N* is not identical with *N**. In these cases, the total amount of pleasure, or pain, that the person will experience is different depending on whether they experience it in the past or future.

In making predictions about conditions of inequality, philosophers have focused mostly on pain. Consider one of the more celebrated thought experiments designed to elicit future-bias: Parfit's (1984, 165) *My Past or Future Operations:*

> 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.

Parfit reports that he would strongly prefer that his operation be in the past, and he predicts that others in the same position would have the same preference.

Strikingly, in *My Past or Future Operations*, this preference is predicted to exist despite the inequality between the payoffs: the past operation involves *ten times* more suffering than the future one. If this is right, then the preference for past suffering over future suffering exists even though it entails a much worse state of affairs from a time-neutral perspective: future-bias here is outweighing the fact that the past surgery involves much more suffering overall. Indeed, Sullivan (2018, 58) has argued that "our discount functions are absolute […] we assign no value to a merely past painful experience or pleasurable experience." Call this *absolute future-bias*. Absolutely future-biased agents prefer a future pleasure to a past pleasure, no matter the magnitude of each, and they prefer a past pain to a future pain, no matter the magnitude of each. By contrast, agents who are non-absolutely future-biased place *some* weight on past pains and pleasures. They simply place *more* weight on future pains and pleasures.

Parfit's thought experiment, and the response in the philosophical literature, shows that philosophers think that people's future-bias is either absolute or, while not absolute, sufficiently robust that they will prefer their suffering be in the past even given a 10:1 inequality. Finding that future-bias can outweigh unequal payoffs in this way would be even more interesting than the discovery that people are future-biased in conditions of equality. After all, in conditions of equality it makes no difference, from a time-neutral perspective, where in time the event occurs: the overall amount of suffering is the same regardless. Perhaps time-neutralists are right, and such a preference in equal conditions is irrational. But it is what we might call *harmlessly irrational*. It doesn't make the people who have those preferences overall worse off. By contrast, time-neutralists argue that the kind of future-bias

that outweighs unequal payoffs can interact with other principles, such as risk-aversion (Dougherty, 2011) or regret-aversion (Greene and Sullivan, 2015), to make people worse off overall. If future-bias is irrational, and these time-neutralists are right, then when future bias outweighs unequal payoffs it is *harmfully irrational*.

Given this, it is important to determine whether the preferences philosophers have attributed to people are descriptively accurate. In part, this is because we want to know whether people's preferences are, if these time-neutralists are right, harmfully irrational. But it is also because many of the arguments put forward in favour of, and against, the rationality of hedonic future-bias are ones that appeal to the descriptive realities of people's preferences. In particular, as we have noted, philosophers often appeal to a supposed asymmetry in our future-bias—between first- and third-person preferences.

Third-person preferences are the preferences that a first-party has about the timing of events for a third-party. Parfit (1984, Section 69) claims that the future-biased preference elicited by *My Past or Future Operations* evaporates if a loved one, and not oneself, either did or will experience the suffering. Hare (2008, 269–70) and Greene and Sullivan (2015, 968–7) both present thought experiments in support of Parfit's claim. Greene and Sullivan's case mirrors the structure of *My Past or Future Operations*. They write:

> Note how our intuitions change when we consider distant people whose circumstances are beyond our control. Suppose your friend from graduate school lives on the other side of the world. You hear through a mutual friend that he requires a painful surgery, which either occurred yesterday and was the longest ever recorded (ten hours) or will occur tomorrow and be much shorter (about an hour). Your informant cannot remember which of these is true. In this case, it is easy to imagine preferring that your friend did not have the longer surgery, even though it would at this point be in his past.

Some time-neutralists, like Brink (2011, 378) and Dougherty (2015, 3), argue that because future-bias is limited in scope it is suspect as a component of a theory of rationality: if we are only future-biased about first-person hedonic events, then that suggests that future-bias is arbitrary and not formed by rational processes.[6]

So philosophers, whether time-neutralists or not, have predicted that people are positively and negatively hedonically future-biased regarding first-person events, and this future-bias often results in them preferring more suffering (or less pleasure) from a time-neutral perspective, but that for third-person events, people are time-neutral.

This paper aims to empirically test these predictions. Doing so is important for at least three reasons. Firstly, time-neutralists appeal to this picture of our preferences to argue for the conclusion that future-bias is irrational. Thus, the descriptive realities of hedonic future-bias play a major role in adjudicating the debate between time-neutralists and defenders of future-bias. Second, if people are indeed future-biased only in first-person hedonic conditions, then whether or not this is rational, it is an interesting and important picture of human psychology and one that, if true, demands further study. Thirdly, if this is the descriptive reality, and if Dougherty (2011) and Greene and Sullivan (2015) are

[6]Greene and Sullivan (2015, Section V) hypothesise that hedonic future-bias is the result of *temporally asymmetric emotions*, which distort our judgment and tend to manifest only with regard to first-person hedonic events. For third-person preferences and non-hedonic first-person preferences, they claim that future-bias does not exist because the party forming the preferences has "emotional distance" from the experiences. Other proposed evolutionary accounts of future-bias suggest something similar. See Horwich (1987, 196–8), Maclaurin and Dyke (2002), and Suhler and Callender (2012).

right about the way in which future-bias interacts with other principles, people may sometimes be significantly worse off because of their future-bias.

In §3 we outline the methodology and results of our study, and in §4 we discuss the upshots of those results for philosophical theorising. First, in §2 we outline extant empirical research in this area and present our predictions. Surprisingly, there has been little empirical work on future-bias, and even less that connects empirical work to the question of its normative status.

2 | EXTANT RESEARCH AND PREDICTIONS

A study performed by Caruso, Gilbert and Wilson (2008) has been influential in the philosophical literature, partly because until very recently it was the only empirical study of future-bias. Caruso et al. asked participants to determine fair compensation for a session of boring data entry work that either occurred in the past or will occur in the future. Participants assigned themselves 60% more compensation for future work versus past work in first-person conditions. However, in third-person conditions, they recommended the same compensation regardless of when the work occurred. This suggested to philosophers that their prediction about a first- versus third-person asymmetry for hedonic future-bias is empirically supported: the participants in the study seemed to be future-biased about their own experiences but time-neutral about the experiences of others.

More recent findings, however, are inconsistent with a straightforward first- versus third-person asymmetry. Greene, Latham, Miller and Norton (2021) showed participants a range of vignettes (a little like those presented by Parfit) and asked them questions about their preferences regarding the locations of events. Like Caruso et al., they found that participants were future-biased regarding first-person positive and negative hedonic events. However, seemingly contrary to the findings of Caruso et al., and to the predictions of Parfit (1984), Hare (2008), Brink (2011), Greene and Sullivan (2015), and Dougherty (2015), Greene et al. did not find time-neutrality when it came to third-person hedonic events. Instead, they found that a significant majority of participants were (positively and negatively) future-biased regarding the location of hedonic events for a third-person.[7]

Importantly, Caruso et al.'s and Greene et al.'s studies asked participants to indicate their preferences regarding whether *the same* hedonic event is past or future: they elicited preferences under what we have called *conditions of equality*. In contrast, Parfit's *My Past or Future Operations* features a payoff of *more* past pain versus *less* future pain. Furthermore, the third-person predictions of Hare (2008) and Greene and Sullivan (2015) seem to exclusively concern unequal payoffs: they predict that people will not prefer a third-party to experience a greater amount of suffering merely because that suffering is in the past. It is not inconsistent with these predictions that participants prefer a third-party's suffering to be in the past in conditions of equality. Indeed, in conditions of equality people might be using the future/past distinction as a tiebreaker to decide between what they consider to be two otherwise equally good states of affairs.

More recent research by Latham, Miller, Norton and Tarsney (2020) suggests that participants are less future-biased in scenarios in which they truly believe they can *affect* the past, than in scenarios in which they truly believe they cannot affect the past. Moreover, participants were less future-biased when the past stakes were larger than the future stakes: that is, in the unequal conditions. However,

[7]Greene et al. (2021) hypothesise that this was because participants treat third-person conditions as first-person conditions when there is sufficient information for them to simulate what they themselves would want in that position (their "simulation hypothesis"). Such information was provided to participants in the Greene et al. study but not in the Caruso et al. study. In the Caruso experiment, participants were asked about an unknown stranger with no personal attributes.

Latham et al. focussed only on negatively valenced hedonic events (namely electric shocks), and only on first-person preferences and choices.

On the basis of the sorts of intuitions voiced by philosophers like Parfit, alongside the empirical work by Caruso et al., Greene et al., and Latham et al., we made the following predictions about the preferences people will have in conditions of inequality. In line with Parfit's case, the particular inequality we chose to test these predictions is 10:1.

Firstly, we predicted that in first-person conditions we would find people's future-bias is strong enough to outweigh the unequal payoffs. In other words, we predicted that, *for themselves,* the mean preference (and the preference of a majority of participants) would be for ten negative hedonic events in the past over one negative hedonic event in the future, and for one positive hedonic event in the future over ten positive hedonic events in the past.

Secondly, combining Greene et al.'s (2021) finding of decreased third-person future-bias (compared to first-person future-bias) in conditions of equality with intuitions from Hare (2008) and Greene and Sullivan (2015), we predicted that in third-person conditions we would find that people's future-bias is not strong enough to outweigh the unequal payoffs. In other words, we predicted that, *for a third-person*, the mean preference (and the preference of a majority of participants) would be for one negative hedonic event in the future over ten negative hedonic events in the past, and for ten positive hedonic events in the past over one positive hedonic event in the future.

3 | EXPERIMENTAL DESIGN AND RESULTS

3.1 | Method

3.1.1 | Participants

522 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.[8] 220 participants had to be excluded for failing to follow task instructions. This means that they failed to answer the questions (126), or failed an attentional check question (94). The remaining sample was composed of 302 participants (aged 22-70; 119 female). Mean age 28.43 (SD = 6.63).[9] 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.

3.1.2 | Materials and procedure

The study was a 2x2 between-participants design, whereby participants were randomly allocated to one of four conditions: first-person positive, first-person negative, third-person positive or third-person negative. In order to allow comparison with Greene et al.'s (2021) data regarding people's preferences

[8]To help ensure the data was collected from human participants disposed to appropriately engage with the experimental materials, each participant was required to have completed at least 1000 surveys with a 95% approval rating.

[9]There were also no significant main effects of age and gender, nor were there any significant interaction effects with age and gender.

under conditions of equality, we used adapted versions of their hedonic vignettes, amended so as to make clear the unequal past and future payoffs. The vignette we used in the first-person positive condition was as follows:

> 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. The ship's dispenser has two different meal-dispensing schedules. On schedule one, it is programmed to dispense your favourite meal *once* during the voyage. On schedule two, it is programmed to dispense your favour meal *ten times* during the voyage. So on schedule two, you receive *ten times* as many favourite meals as you do on schedule one.
>
> One morning, you awake from a dream concerning your favourite meal and for a moment you cannot remember which schedule your dispenser is programmed with. If it is programmed with schedule two, then you received your favourite meal yesterday, and nine times earlier in the voyage. If it is programmed with schedule one, then you will receive your favourite meal tomorrow.

After reading this vignette, participants were presented with *one* of the following statements (which statement each participant saw was randomised), to which they responded on a Likert scale from 1 (strongly disagree) to 7 (strongly agree).

(a) I would prefer to learn that I will receive my favourite meal only tomorrow (instead of yesterday and nine times earlier in the voyage).
(b) I would prefer to learn that I received my favourite meal yesterday and nine times earlier in the voyage (instead of only tomorrow).

The vignette we used in the first-person negative condition was as follows:

> 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. The ship's dispenser has two different meal-dispensing schedules. On schedule one, it is programmed to dispense your most disliked meal (which you really dislike) *once* during the voyage. On schedule two, it is programmed to dispense your most disliked meal *ten times* during the voyage. So on schedule two, you receive ten times as many of your most disliked meals as you do on schedule one.
>
> One morning, you awake from a dream concerning your most disliked meal and for a moment you cannot remember which schedule your dispenser is programmed with. If it is programmed with schedule two, then you received your most disliked meal yesterday, and nine times earlier in the voyage. If it is programmed with schedule one, then you will receive your most disliked meal tomorrow.

As in the positive condition, participants were then presented with *one* of the following statements (which statement each participant saw was randomised), to which they responded on a Likert scale from 1 (strongly disagree) to 7 (strongly agree).

TABLE 1 Descriptive data from all conditions.

	%Yes	%No	%4	Mean	SD	t-test	*p*-value	χ^2	*p*-value
Condition 1: First-Person Positive (N = 70)	12.9	65.7	21.4	3.11	1.57	−4.734	<.001	6.914	.009
Condition 2: First-Person Negative (N = 88)	62.5	18.2	19.3	4.67	1.44	4.355	<.001	5.500	.019
Condition 3: Third-Person Positive (N = 74)	16.2	64.9	18.9	3.27	1.41	−4.460	<.001	6.541	.011
Condition 4: Third-Person Negative (N = 70)	77.1	12.9	10.0	5.14	1.61	5.944	<.001	20.629	<.001

a) I would prefer to learn that I receive my most disliked meal only tomorrow (instead of both yesterday and nine times earlier in the voyage).
b) I would prefer to learn that I received my most disliked meal yesterday and nine times earlier in the voyage (instead of only tomorrow).

Participants in the third-person conditions saw the same vignettes and responded to the same statements, but each was systematically amended with third-person locutions. Thus each vignette begins "Freddie is an astronaut on a 10-year voyage between planets…", and, for example, statement (a) for the third-person positive condition was "I would prefer to learn that Freddie will receive his favourite meal only tomorrow (instead of both yesterday and nine times earlier in the voyage)."

The reason for presenting half of the participants in a given condition with statement (a) and half with statement (b) was in order to control for question effects, in particular acquiescence bias, whereby people are generally more inclined to agree with statements than to disagree with them. In all four conditions, participants were asked an attentional check question: "In the vignette you were asked to read, on which schedule [do you]/[does Freddie] obtain *more* of [your]/[his[[favourite meal]/[most disliked meal] in total?" Participants had 4 response options (schedule one; schedule two; schedule three; schedule four) and those who failed to answer the question correctly were excluded from the study.

In order to amalgamate these results in the positive conditions, levels of agreement with the latter kind of statement were reverse-coded (i.e., a response of 1 was transformed into a response of 7, a response of 2 was transformed into a response of 6, and so on). In the negative conditions, levels of agreement with the former kind of statement (expressing a preference for the single most disliked meal in the future) were reverse-coded. After this reverse-coding, the results are *as if* all participants had been asked, in the positive conditions, whether they would prefer a single positive future event to ten positive past events, and in the negative conditions, whether they would prefer ten negative past events to one negative future event. Thus, in what follows, *higher levels of agreement* indicate *stronger future-biased preferences*.[10]

3.2 | Results

Table 1 and Figure 1 below summarise the descriptive data from the experiment. After the reverse-coding described in §3.1.2, the 'yes' column represents the proportion of participants who agree with the statement that they have future-biased preferences and the 'no' column represents the proportion of participants who *disagree* with the statement that they have future-biased preferences. The '4' column represents the proportion of people who neither agree nor disagree with those statements. We also included the results of the t-tests and one-way χ^2-tests we ran for each condition.

The results of our t-tests show mean levels of agreement significantly above 4 in the two negative conditions ((2) and (4)), and significantly below 4 in the two positive conditions ((1) and (3)). Thus, we find future-biased preferences in the negative conditions—on average people prefer ten negative events in the past to one negative event in the future, both for themselves and for Freddie—and non-future-biased preferences in the positive conditions—on average, people prefer ten positive events in the past to one positive event in the future, both for themselves and for Freddie.

[10]There was no significant main effect of statement type, nor were there any significant interaction effects with statement type. Statement type has *no influence* on our reported results. As such it is permissible to amalgamate the two statement types in the manner that we have.

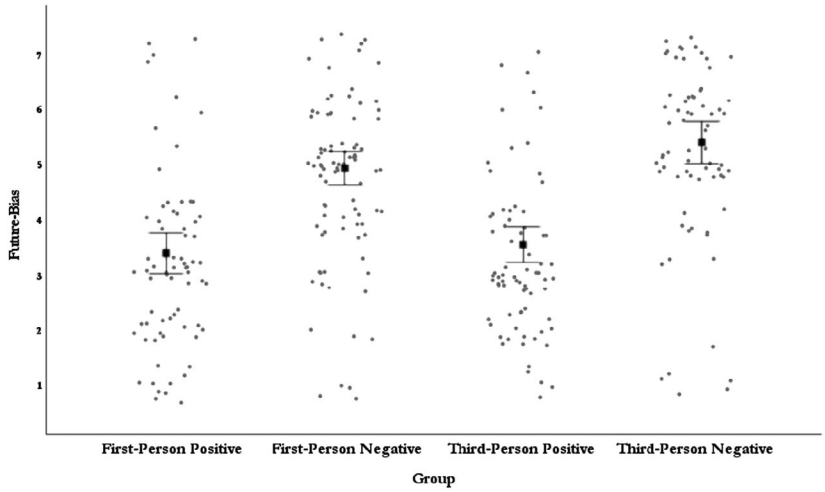

FIGURE 1 Distribution of participant responses in all conditions. Mean point and error bars reflect 95% confidence interval. Points are jittered to avoid complete overlap.

However, a mean significantly above 4 can come about despite most participants responding with 4. (Likewise, a mean that does not differ significantly from 4 can come about not as a result of most participants responding with 4, but as a result of there being a balance between responses above and below 4.) Thus, it is perhaps more illuminating to consider the results of our one-way χ^2-tests.

In the two negative conditions, where the mean was significantly above 4, the χ^2-tests show us that a significant majority of participants responded with either 5, 6, or 7. In other words, a significant majority of participants preferred ten negative past events to one negative future event, both for themselves and for Freddie.

In the two positive conditions, where the mean was significantly below 4, the χ^2-tests show us that a significant majority of participants responded with either 1, 2, or 3. In other words, a significant majority of participants preferred ten positive past events to one positive future event, both for themselves and for Freddie.

To compare strength of future-biased preferences across conditions, we tested level of agreement with a 2x2 between-subjects ANOVA. There was only a main effect of valence, such that future-biased preferences were stronger in negative conditions ($M = 4.91$, $SD = 1.51$) than in positive conditions ($M = 3.19$, $SD = 1.50$), $F(1, 298) = 97.295$, $p < .001$.

4 | DISCUSSION

We predicted, on the basis of the philosophical and empirical literature, that participants' future-bias about first-person experiences would be strong enough to outweigh the 10:1 inequality, and thus that participants would report future-biased preferences in both positive and negative first-person conditions. By contrast, we predicted that participants' future-bias about third-person experiences would not be strong enough to outweigh the 10:1 inequality, and thus that participants would report non-future-biased preferences in both positive and negative third-person conditions.

We did not, however, find the predicted asymmetry between first- and third-person conditions. Instead, we found symmetry between first- and third-person conditions, and asymmetry between positive and negative conditions.

More specifically, we found future-biased preferences in both negative conditions. That is, participants preferred, both in terms of mean response and in terms of the majority of participants, ten negative events in the past to one negative event in the future (both for themselves and for Freddie). This is what we predicted in the first-person negative condition, but not in the third-person negative condition. This finding is interesting, since the discovery of future-biased preferences in these conditions is good evidence of the existence of first- and third-person negative hedonic future-bias that is not outweighed by unequal payoffs.

In contrast, we did not find future-biased preferences in either positive condition. This is what we predicted in the third-person positive condition, but not in the first-person positive condition. In these conditions, we found that both in terms of the mean response and in terms of the majority of participants, people preferred ten positive past events to one positive future event (both for themselves and for Freddie).

We thus have strong evidence against Sullivan's (2018) hypothesis that people are absolutely future-biased. It seems that people do assign value to past positive hedonic events. While it is most natural to interpret this pattern of responses as an indication that the majority of participants are non-absolutely future-biased about positive hedonic events, we should note that the lack of future-biased preferences in the positive conditions could also be explained by participants being *past*-biased about positive events: they might just prefer positive events to be past rather than future. Likewise, these results are consistent with participants having time-neutral preferences regarding positive events, and simply preferring the state of affairs that, from a time-neutral perspective, maximises utility.

Previous research, however, can help us rule out past-bias and time-neutrality. We know that in conditions of equality people do not show past-biased or time-neutral preferences about first- or third-person positive hedonic events (Greene et al. (2021); Greene, Latham, Miller and Norton (forthcoming)). So we have no reason to think that people would be past-biased or time-neutral in conditions of inequality. We thus conclude that in the positive conditions the 10:1 ratio was enough to outweigh a non-absolute future-bias.[11]

Indeed, if we compare our results in the 10:1 condition to the results of Greene et al. (2021), it appears as though people are less future-biased in the 10:1 condition than in the 1:1 condition of the Greene study. This suggests that participants have a non-absolute future-bias regarding both positive and negative events. Drawing together first- and third-person responses, Greene et al. reported that in equal conditions, the mean response was 5.42 (*SD* 1.62) for negative events, while in unequal conditions we found a mean of 4.91 (*SD* 1.63). This modest decrease suggests that people's preference to have pains in the past rather than the future is attenuated when the past pain is greater than the future pain. More substantially, while Greene et al. reported that in equal conditions, the mean response was 5.21 (*SD* 1.62) for positive events, in unequal conditions we found a mean of only 3.19 (*SD* 1.62). Of course, we cannot determine whether the apparent decrease in future-bias in unequal versus equal conditions is statistically significant. This is somewhere where future work could profitably be undertaken. If the appearances are vindicated, however, then these results would show that the presence of the inequality partially mitigates people's future-biased preferences regarding both positive and negative events, and hence lend further weight to the conclusion that they are not absolutely future-biased in either condition.

[11]While we cannot rule out the possibility that participants are time-neutral about positive events in unequal conditions, this hypothesis would require one to posit that people shift from a future-biased to a time-neutral perspective when considering positive experiences instead of negative experiences, and when considering unequal positive payoffs instead of equal positive payoffs (as Greene et al. (2021) found participants to be future-biased for positive equal payoffs). Thus, one would have to think that people are time-neutral in only one hedonic category—positive unequal payoffs—and future-biased otherwise.

In sum, we believe our results are best explained by participants having a non-absolute future-bias in both positive and negative conditions. Future research that tests for hedonic future-bias at different inequalities would be instructive here (more on this below).

The two main upshots of this research, therefore, are the following: i) there is no robust asymmetry in future-bias between first- and third-person preferences, and ii) negative future-bias is stronger than positive future-bias. These upshots have important implications for normative debates about the rationality of future-bias.

First, recall that Greene et al. (2021) also failed to find a first- versus third-person asymmetry in conditions of equality: participants were future-biased in both first- and third-person equal hedonic conditions. However, they did find that future-biased preferences were significantly stronger in first-person conditions than in third-person conditions. By contrast, in conditions of inequality, we did not find any kind of significant difference between first- and third-person conditions.

This result tends to undermine time-neutralist arguments that hedonic future-bias is irrational because it is inconsistently applied to first- and third-person preferences. Our findings strongly suggest that people's preferences do not display the robust asymmetry attributed to them by time-neutralists. Insofar as there is a difference between first- and third-person preferences, it is most obviously present in conditions of equality, but even there the majority of participants had future-biased preferences in both first- and third-person conditions. Thus, this argument is unsuccessful. This is not to say, of course, that future-bias is rational: it is only that one sort of argument for time-neutralism is founded on an empirically suspect claim.

Consider now our second upshot: future-bias is strong enough to outweigh a 10:1 inequality only in the negative conditions. As we noted above, philosophers who make predictions about future-bias in unequal conditions have focused on negative events like painful experiences. Supporters of the rationality of future-bias, in particular, almost exclusively appeal to cases involving future or past pain.[12] Our results offer an explanation for this practice: the strength of future-bias is greater for negative events.

On the basis of their results concerning conditions of equality, Greene et al. (2021) hypothesised that future-bias tends to increase when events are negatively valenced. However, the increase in future-bias that Greene et al. (2021) observed was small in comparison to the increase observed in the present study. This suggests that the change from equal to unequal payoffs has served to reveal a more substantial difference between positive and negative future-bias. The best explanation for this, given the informal comparison of our results with those of Greene et al. (2021) discussed above, is that people are non-absolutely future-biased with regard to positive events, and perhaps also with regard to negative events. Further, the strength of their future-bias is much weaker for positive events. Thus, people's preferences regarding equal payoffs look similar between positive and negative events because people's future-bias is determining their preference regarding both kinds of events. By contrast, people's preferences regarding 10:1 unequal payoffs look different between positive and negative events because it is the interaction between future-bias and unequal payoffs that is determining their preferences, and the inequality is enough to outweigh positive future-bias but not negative future-bias. This suggests that Greene et al.'s (2021) results were masking a substantial difference between our temporal preferences regarding positive and negative events.

This newfound asymmetry opens up a new avenue for normative arguments in favour of time-neutralism, which have a similar structure to the argument that appeals to the alleged first- and third-person asymmetry undermined by our results. In the absence of an account of why we ought to treat positive and negative events differently—in particular, why we ought be more future-biased regarding

[12]See, e.g., Prior (1959), Hare (2007, 2008) and Heathwood (2008).

the latter than the former—these results undermine the claim that future-bias is a rational reaction to a metaphysical difference between past and future experiences.

Further work would be valuable testing ratios smaller than 10:1, to establish the 'tipping point' at which positive future-bias is outweighed. Given *five* favourite meals in the past versus *one* in the future, would participants' future-bias be outweighed? What about *two* favourite meals? We don't know, and we hope to take this up in future work.

Ultimately, our results show that the future-bias for negative experiences reported by Greene et al. (2021) was not the result of a mild or 'tiebreaker' preference that was reported only because all else was equal. Instead, it seems that negative hedonic future-bias is strong enough that people prefer things to be *much worse* overall in order to have negative events in their past. Once again, future work would be valuable in investigating the limits of these preferences. If negative future-bias is non-absolute, for instance, then there is some inequality, greater than 10:1, at which it will be outweighed. What ratio this would be, we don't know, and once again we hope to take this up in the future.

What we certainly can say is that both time-neutralists and defenders of future-bias are right to think that there are kinds of preferences—namely, preferences about the temporal location of negative events—that display future-bias even in conditions of inequality. If time-neutralists are right and these preferences are irrational, then people are irrational in a non-trivial way: they are preferring that overall, things are worse. In addition, if the arguments of Dougherty (2011) or Greene and Sullivan (2015) succeed, these preferences are *harmfully irrational*: they make people's lives go worse.

5 | CONCLUSION

The results of our study show that philosophers' general prediction that people sometimes prefer more pain in the past to less pain in the future is correct. However, in other ways our results were contrary to what many philosophers have predicted. We found that future-bias was strong enough to outweigh unequal payoffs only for negative hedonic events, at least at the 10:1 ratio we tested. For positive hedonic events, people preferred the state of affairs that is best overall (i.e., best from a time-neutral perspective). Our study also found no asymmetry between people's first- and third-person preferences in conditions of inequality.

The lack of a first- versus third-person asymmetry undermines a popular time-neutralist argument against the rationality of hedonic future-bias. However, the asymmetry in future-bias that we observed between positive and negative experiences provides a potential replacement argument. Defenders of the rationality of future-bias almost exclusively focus on thought experiments comparing more past pain to less future pain, and this seems to be for good reason, as our study shows. But why is future-bias attenuated in the case of positive experiences? The fact that people's future-bias responds differently to positive and negative experiences may suggest that future-bias is either arbitrary or the result of evolutionarily instilled asymmetric emotions.[13]

REFERENCES

Brink, D. O. (2011). Prospects for temporal neutrality. In C. Callender (Ed.), *The Oxford handbook of philosophy of time* (pp. 353–381). Oxford: Oxford University Press.

[13]We are grateful to several anonymous referees for constructive comments on earlier drafts. Kristie Miller would like to thank the Australian Research Council (FT170100262 and DP18010010); James Norton would like to thank the Icelandic Centre for Research (195617-051); Andrew J. Latham would like to thank the Ngāi Tai Ki Tāmaki Tribal Trust; and Preston Greene would like to thank the Singapore Ministry of Education Academic Research Fund Tier 1 RG134/19(NS) for their support.

Caruso, E., Gilbert, D. T., & Wilson, T. D. (2008). A wrinkle in time: Asymmetric valuation of past and future events. *Psychological Science*, *19*(8), 796–801.

Dougherty, T. (2011). On whether to prefer pain to pass. *Ethics*, *121*(3), 521–537.

Dougherty, T. (2015). Future-bias and practical reason. *Philosophers' Imprint 15*(30), 1–16.

Greene, P., & Sullivan, M. (2015). Against time bias. *Ethics*, *125*(5), 947–970.

Greene, P., Latham, A. J., Miller, K., & Norton, J. (2021). Hedonic and non-hedonic bias towards the future. *Australasian Journal of Philosophy*, *99*(1), 148–163.

Greene, P., Latham, A. J., Miller, K., & Norton, J. (forthcoming). Why are people so darn past biased? In C. Hoerl, T. McCormack, & A. Fernandes (Eds), *Temporal asymmetries in philosophy and psychology*. Oxford: Oxford University Press.

Hare, C. (2007). Self-bias, time-bias, and the metaphysics of the self and time. *Journal of Philosophy*, *104*(7), 350–373.

Hare, C. (2008). A puzzle about other-directed time-bias. *Australasian Journal of Philosophy*, *86*(2), 269–277.

Hare, C. (2013). Time—The emotional asymmetry. In H. Dyke & A. Bardon (Eds.), *A companion to the philosophy of time* (pp. 507–520). West Sussex: Wiley-Blackwell.

Heathwood, C. (2008). Fitting attitudes and welfare. In *Oxford studies in metaethics* (Vol. *3*, pp. 47–73). Oxford: Oxford University Press.

Hedden, B. (2015). *Reasons without persons: Rationality, identity, and time*. Oxford: Oxford University Press.

Horwich, P. (1987). *Asymmetries in time: Problems in the philosophy of science*. Cambridge: MIT Press.

Kauppinen, A. (2018). Agency, experience, and future bias. *Thought: A Journal of Philosophy*, *7*(4), 237–245.

Latham, A. J., Miller, K., Norton, J., & Tarsney, C. (2020). Future bias in action: Does the past matter more when you can affect it? *Synthese*. https://doi.org/10.1007/s11229-020-02791-0

Maclaurin, J., & Dyke, H. (2002). 'Thank goodness that's over': The evolutionary story. *Ratio*, *15*(3), 276–292.

Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.

Prior, A. N. (1959). Thank goodness that's over. *Philosophy*, *34*(128), 12–17.

Suhler, C., & Callender, C. (2012). Thank goodness that argument is over: Explaining the temporal value asymmetry. *Philosophers' Imprint*, *12*, 1–16.

Sullivan, M. (2018). *Time biases*. Oxford: Oxford University Press.