Australasian Philosophical Review

Volume 5 Issue 3 September 2021

Routledge
Taylor & Francis Group

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rapr20

Response to Critics

Craig Callender

To cite this article: Craig Callender (2021) Response to Critics, Australasian Philosophical Review, 5:3, 309-321, DOI: 10.1080/24740500.2021.2112131

To link to this article: https://doi.org/10.1080/24740500.2021.2112131

Published online: 09 Feb 2023.

AUSTRALASIAN PHILOSOPHICAL REVIEW
2021, VOL. 5, NO. 3, 309–321
https://doi.org/10.1080/24740500.2021.2112131

Response to Critics

Craig Callender

University of California, San Diego

Let me begin by sincerely thanking the commentators for taking the time to share their insightful reactions to the target article [Callender 2022a]. I am very fortunate to have so many talented, diverse and rich responses. They have put me on an intellectual journey for the past two years, one that has been very rewarding and led me to change and hopefully improve my view. These developments wouldn't have been possible without such challenging feedback.

To avoid missing the forest for the trees, let me highlight what is most important to me. It's a practical problem. In some social sciences and especially cost-benefit analysis it is common to blame many problems on people having 'irrational' time preferences. In neo-classical economics the value of the temporal discount rate was often targeted as the cause of many individual and societal ills. People discounted future value too steeply, it was thought, not caring about the future. Modern researchers instead focus on the form of the temporal discount function. Due largely to a dominance result by Strotz [1955], an exponential form is considered the normatively correct function. Non-exponential functions, which it turns out the folk use, are thought to be temporally inconsistent and therefore irrational. Because credit card debt, obesity, smoking, and more are associated with non-exponential discounting, our 'irrational' time preferences are thought to partly cause our misfortunes, which in turn suggests certain types of policy intervention.

Not everyone subscribes to this paradigm—but most do.[1] Yet this paradigm often gets things the wrong way around. In neo-classical economics, the poor were blamed for their fate because it was said that they didn't care about the future. But, of course, the causal arrow might go the other way: when food and shelter are under threat and mortality high, it can be entirely rational to not care so much about the distant future. Modern work that focuses on the form of the discount function faces the same possible reversal. Given certain contexts, it may make sense to employ non-exponential discount functions. What I try to show is that the number of assumptions needed to bestow a normative glow on exponential discounting are substantial; without getting into the dirty contextual details, we cannot assume that non-exponential discounting is irrational [Callender 2022a].

The standard paradigm is quick to assume time preferences are the problem. This style of thinking can be traced right back to Strotz's seminal paper. He writes

> Spendthriftiness, in the general sense of inconsistent or imprudent planning, is by no means insignificant. It is especially among the lower-income classes, where education and training

CONTACT Craig Callender ccallender@ucsd.edu

[1] See Becker and Murphy [1988] for a quite different paradigm, one famously used to account for addiction.

> are commonly blighted, that one would expect to find imprudent behaviour of this sort. In America, lower-income people tend to gorge themselves with food after pay day; overheat their homes when they have money for a bucket of coal; are extravagant, going on sprees on pay-day, not budgeting their money, and engaging in heavy instalment buying; do not keep their children in school, and are freer in the expression of their sexual and aggressive impulses. Their high birth rate is well-known. All these behaviour characteristics can be explained as a failure to cope intelligently with the problem of the intertemporal tussle [ibid.: 179].

Apart from the focus on temporal inconsistency rather than high discount rates, this passage sounds like something said in the late 19th-century.[2] Strotz [ibid.: 178n5] contemplates the opposite causal direction, that temporally inconsistent behaviour patterns might be due to insecurity. But 'it is not clear' to him why this should be the case. This failure of imagination speaks volumes.

Here is a recent example based on the association between BMI (body weight index) and hyperbolic discounting. Going against the conventional wisdom, Richards, Hamilton, and Pofahl [2010] posit that high BMI causes hyperbolic discounting due to obesity's effect on mortality. While acknowledging that actually showing it would require a lot of work, Ikeda [2016: 137] dismisses this argument and claims the reverse causal direction is intuitively more plausible. The obese rack up a 'fat debt' [ibid.: 135] for the same alleged reason the poor suffer credit card debt: poor time preference. Clearly the causal situation here is tremendously complex, given that the obese tend to be poor, the poor have less healthy food options, greater risks, more stigma, and more—to say nothing of the effect of genetics. What makes Ikeda's favoured causal direction so intuitive is the standard normative template.

What does the causal direction have to do with the normative standard? The world's causal network is what it is, independent of the normative standard. However, because the standard normative paradigm views hyperbolic discounting as nearly always a *mistake*, the default is to blame these preferences as the cause of misfortune. This template has become standard in the field, although there are notable exceptions. In Strotz's examples, clearly the causal influences amongst all these variables (education, income, number of children, time preference, and so on) are hardly obvious. What makes him and others confidant that time preference is to blame? Answer: this default paradigm. Where this is especially worrying is when policy is then suggested, as it tends to favour policies that shift burdens on to the socially vulnerable.

My [Callender 2022a; see also Callender 2023] describes good reasons why one might discount non-exponentially. To get Strotz's result, we need two key assumptions, Consistency and Invariance. Consistency holds that as you move through time you continue to honour your earlier self's preferences. Invariance holds that you keep the same preferences as the world changes. Consistency faces the problem of changing selves. Invariance faces the problem of drawing a distinction between pure and impure temporal preferences. Before we accuse people of making a mistake, we need to be confident that these two conditions hold—and typically we cannot be.

None of the commentators, I think, object to this real-world concern. But my worry is compatible with the normative standard being correct but misapplied. It is also compatible with a new normative standard. If we ditch exponential temporal

[2] Disturbingly, Strotz [1955: 178n4] cites approvingly Fisher's claim that sharp discounting was more common among 'primitive races, children, and other uninstructive groups in society'.

discounting as the normative standard, we need a new one. Steele and Grüne-Yanoff each offer conservative modifications of the existing normative standard. Greene, as we'll see, wants something even stricter, no discounting at all, exponential or not. Robertson suggests enactive rationality. The 'bounded rationality' position is another natural reaction, as we'll see when discussing Smartt's reply. I'm not ready to embrace any just yet. Right now, I want to assemble and study the menu. How should we temporally discount, if at all?

Katie Steele

Because I will later draw on my reaction to Steele's [2022] comments, let me take them first. Like me, Steele thinks finding a clean separation between pure and impure temporal preferences isn't a promising path forward. Like Grüne-Yanoff, she is worried about flip-flopping (preference reversals). She suggests a way to keep the focus on flip-flopping without having to draw a line between pure and impure (tenseless) temporal preferences. Moreover, she argues, correctly in my opinion, that this is what Strotz originally intended too. In what follows, I'll describe an interesting way of looking at this important move.

Steel draws our attention to that fact that Strotz distinguishes two types of discounting. One is discounting from the present moment. This is represented in the discount function D. The other is discounting from the calendar date of the future consumption. The second are 'impure' temporal preferences. He gives an example of wanting champagne on his birthday; receiving it later isn't as valuable to him. This temporal preference gets represented in the utility function and not the discount function. The weight assigned to the pleasure of champagne, say $\boldsymbol{x}$, becomes a function of calendar time t, $u(\boldsymbol{x},t)$, where the value is high when t = birthday and lower afterwards.

Now we don't have preferences about champagne delivered at different times, we have preferences between champagne-on-March 15 and champagne-on-March 16. They're like two different types of champagne. I value champagne-on-March 15 but not the other one because I have the 'impure' information that March 15th is my birthday. But this I know and plan for all along. With the 'impure' built into the utility weights, I may be able to model the decision-maker as satisfying Invariance. If so, then I need check on Consistency. It might be that I prefer the early champagne at all evaluation times, i.e., for all values of τ, where τ is the time of evaluation. If so then there is no flip-flopping.

Steel's point was a revelation to me. Being a philosopher of time, the distinction between tensed and tenseless time series is always in mind. Following McTaggart [1908], a tensed time series locates events in the past, present or future; a tenseless time series locates events in sequential or calendar time, a timeline organized by the earlier than relation. When I began working on discounting, I had the nagging worry that this distinction is relevant. I immediately noticed, for instance, that although actual discounting is tensed—delay discounting is delay from the present evaluation point or Now, not from a tenseless calendar time—the normative standard is tenseless. That is, *when is Now* doesn't matter to exponential discounting: it says you should discount proportional to calendar time. The Now drops out [Callender 2022b]. In addition, I saw that tenseless discounting could be consistent even if non-exponential. Suppose I decide to discount goods in the 2020's at 3%, goods in the 2030's at 4%, and goods in the 2040's at 7%. So long

as I hold true to this, I am not susceptible to the preference counterpart of a Dutch book.[3]

When I understood Steele's point everything snapped into place for me. I saw that Invariance is essentially the statement that your preferences should be invariant across calendar time and that Consistency is essentially the statement that your preferences should be invariant across changes of tensed perspective, i.e., past, present, future. Her proposed new normative standard—Received View*—is then the claim that your discounting should not hang on the time of evaluation point (if you can be modelled as satisfying Invariance). Received View* isolates the nasty kind of flip-flopping that one worries about when evaluating plans. This is a core thought behind temporal neutrality. I'm not sure if she would put it this way, but if we conceive of the time of evaluation as indexical in nature, then the idea is that in so far as one is rational one's temporal preferences do not depend on temporal indexicals.

Received View* represents a significant advance. Note that her position is compatible with what I care about. Consider the association between BMI and hyperbolic discounting. Is this a signal that the obese have irrational time preferences? Not necessarily. Many or most could be like Steel's example of Margot, who values oysters on vacation week but not otherwise—they are a holiday thing for Margot. Similarly, if a person can't afford healthy meals, their choices may be perfectly rational and indicate no inconsistent planning. So on Steel's position, I can keep what I care about, isolate nasty temporal flip-flopping, and not have to draw a line between pure and impure temporal preferences. I can get behind this.

The problem of changing selves is still an issue. And there are other problems with the discounted utility model (see e.g., Bleichrodt and Gafni [1995]). For contexts where we have stable preferences and these other problems aren't relevant, Received View* strikes me as an attractive position.

Till Grüne-Yanoff

Grüne-Yanoff [2022] also doesn't like flip-flopping. Not forming self-sabotaging beliefs is a necessary condition of planning, he says. Whether the plan works out for good or ill, the possibility of making a plan is an important part of being rational. For this reason, like Steele, he wants to defend the normative force of Consistency but not Invariance. Since Invariance is needed to derive Stationarity, he is not interested in defending the exponential form of discounting as uniquely rational. To the extent that one discounts hyperbolically because one violates Consistency, however, he would find that type of hyperbolic discounting sub-optimal.

Grüne-Yanoff builds a compelling case for Consistency, and I now wish that I had better marked its virtues and circumscribed where I think it does have normative bite. I agree with many of his counter-arguments to mine, but I like to think that we're not really so far apart if we bracket a larger background issue. In the more formal areas of epistemology belief reversal is considered the major sin. The possibility of a reversal means that one can be exploited by a Dutch book. Knowingly adopting a strategy that can possibly lead to such exploitation seems self-sabotaging and irrational. This verdict is far and away the dominant view in these fields. Influenced by Foley [1993]

[3] On consistent non-exponential discounting, let me highlight Harvey [1994], which is excellent but I failed to use in the target article.

and others, however, I've never fully subscribed the idea that the impossibility of being Dutch booked was either necessary or sufficient for being rational.

Strotz's result is about preference reversals, but the dialectic here is the same as with beliefs. My paper [Callender 2022a] rehearses some common objections to using belief reversals as a criterion of rationality, but now transferred to the case of preference reversals. Grüne-Yanoff then rehearses some natural and reasonable replies to these worries. But this debate is bigger than temporal preferences and I mostly agree with him. I do want to make room for the possibility of some violations of Consistency being epistemologically blameless or innocent—some changes of taste can be rational. Yet I can agree with Grüne-Yanoff that many violations of Consistency are epistemically blameworthy. Someone who can't hold a job because their temporal discount rate is sabotaging their use of an alarm clock has a big problem.

Bracketing that disagreement, I suspect Grüne-Yanoff would like the compromise position sketched by Steele. Like Steele, Grüne-Yanoff thinks the flip-flopper who violates Consistency is doing something wrong. All that 'tenseless' temporal discounting associated with Invariance doesn't threaten Consistency, so Steele's position would paint the target on Consistency, just as Grüne-Yanoff wants. Accepting this, he wouldn't have the burden of defending the pure versus impure distinction. As I understand things, that is a problem for Invariance, not Consistency. But he would have to say something about the problem of changing selves, which I see as the main challenge to Consistency. I don't see a response to this problem in his reply. If we assume that we're in situations where our selves aren't changing much, however, then our positions are not so far apart.

To be clear, I agree with his conclusion that 'there are good reasons to judge certain interactions between time and preference as *generally* irrational, *ceteris paribus*' [Grüne-Yanoff 2022: 275]. And I disagree with his characterization of my view as being 'a matter of contextual dependence and subjective tastes' [ibid.]. That is not my view. What is going on is that we are emphasizing different aspects of what is going on. Look again at his conclusion. There is a tension between 'generally' and 'ceteris paribus.' That nicely represents our disagreement, for he is highlighting the 'generally' (often Consistency has normative bite) and I am the 'ceteris paribus' (sometimes it doesn't).

As the BMI case highlights, it can be very hard to tell if the ceteris paribus conditions hold. Why does the link between high BMI and hyperbolic discounting exist? Some people with high BMI might be consistent planners but discount non-exponentially just because they are poor and can't afford healthy food. Perhaps they violate Invariance but not Consistency. Others will violate Consistency. Some may violate it because they hold self-sabotaging time preferences. I agree with Grüne-Yanoff that these flip-floppers may be irrational. Others may violate it due to endogenous preferences developed in response to an economic shock. No doubt the real story is very complex. Janssens, Kramer and Swart [2017] was a great motivation to me. When you actually do the temporal discounting experiment over two evaluation times, one finds many epistemically 'innocent' reasons for violating the conventional normative standard.

Ian Robertson

Robertson [2022] does not contest my attack on the standard model of temporal discounting. Rather, he criticizes my discussion of dual-system approaches to temporal discounting and then suggests that I replace the conventional normative standard with a revisionary understanding of rationality.

I mention dual-system theories only to highlight that the standard paradigm really is a full 'soup-to-nuts' meal. It comes with results in high theory regarding how we ought to discount, a ton of experimental data, and even a mechanism that is supposed to reinforce the normative ideal, namely, dual-system theory. The idea is that we have two different cognitive systems, one that is fast, automatic and unconscious (System 1) and another that is slow, reflective and conscious (System 2). I like Robertson's quote from Grayot [2020: 124] characterizing the picture: System 2 'corresponds to an inner rational agent, one that would otherwise abide by the norms of expected utility theory were it not for the failures of System 1'. Here the idea is that impulsive System 1 makes us hyperbolic rather than exponential discounters.

Robertson's objection to my treatment is that I bounce off only one kind of dual-system theory, Metcalfe and Mischel [1999]. I make some objections to the idea that their model explains the hyperbolic vs exponential distinction. However, some of my objections are too quick and target features peculiar to the Metcalfe-Mischel model. Had I considered other more standard models, Robertson says, some of my objections would not have landed. I agree.

Not only do I agree with his criticism, but I thank Robertson for his survey of the wide variety of dual-system models. I learned a lot.

The upshot of this survey does seem to be that my main point is strengthened. Draw up the divide between System 1 and System 2 systems in any reasonable way—it will not map neatly onto the exponential versus hyperbolic divide. Robertson [2022: 299] tells us that the picture described in the Grayot quote above is being 'gradually overturned.' Cognitive scientists are not finding two independent forms of thinking that underlie the divide between rational and irrational thinking. This point improves and extends my worries about dual-system explanations of temporal discounting. Dual-system theories are simply too unclear to provide additional confirmatory support for what I dubbed the Standard Model. This is an important point because this support helps prop up and motivate the standard paradigm.

Robertson concludes with an invitation for me to consider more radical conceptions of rationality, in particular, conceptions coming from the enactive cognitive science tradition. On these conceptions rationality is understood as proficient interactions by an organism with its environment. I wonder what exactly 'proficient' means. Presumably some adaptive interactions are more proficient than others, but in my limited acquaintance with this theory, I don't yet see what provides the standard. I'm probably not ready for anything so radical, but I'll investigate.

I can see why Robertson raises this suggestion. If we step back, I'm pushing a more 'public health' oriented look at temporal discounting as an alternative to the normal 'economic' picture. The economist suggests that hyperbolic temporal preferences are causing all sorts of problems. Instead, I'm asking why people are in conditions such that hyperbolic discounting is rational, looking at their behaviour as it is shaped by their environmental context. This outlook is very congenial to enactivism because the environment is a crucial feature of both approaches.

Kristie Miller, James Norton and Andrew Latham

Miller et al. [2022: 282] claim that their 'empirical research serves as a vigorous thump on the table' in support of my argument. As a huge fan of Miller and her colleagues' research in philosophy of time —and especially Miller's recent empirical work on

past time preferences—I am delighted to receive such support. I agree that these 'vigorous thumps' do substantially add to the points I make. Her point about entropy, deliberation and impurity is a deep one, and her research revealing a surprising diversity in past-directed time preferences amplifies my point that finding the normative in this sea of possibly confounding variables is harder than most acknowledge.

Since we're in mutual agreement, let me use my space to comment on past discounting. Miller's research group has done more to advance this sub-field than anyone. While there have been hundreds or thousands of empirical studies of future discounting, there has been only a handful on past discounting, most by her group. Miller's work raises many interesting questions about past discounting. Does past discounting predict anything about future discounting? Her studies suggest the answer is no. She asks: is the reason for this because past discounting is 'more pure' than future discounting? That's an interesting idea, one I should have but never really considered. Let's pursue this thought.

In the future direction, our preferences are being modulated by factors such as uncertainty, mortality, and psychological states like anticipation. The past controls for an awful lot of that. Impulsivity seems not to be an issue. Since 'I' flow forward in time and the rewards are hypothetical, I don't have to wait longer for a reward delivered in the distant past than in the near past. I don't have to wait at all. And I know I survived my past, so mortality is not an issue either. The past doesn't control for everything, of course. I can still be uncertain about past events. Flip a coin but cover it up: I still don't know if it's heads or tails. And future impurities can sneak into past-directed preferences. I might wish that my PhD happened closer to the present so I could apply for junior scholar grants tomorrow. But, overall, perhaps past-directed preferences are purer.

Here is another reason to think past discounting might be more pure. Temporal discounting can include a perspectival shift and a metrical shift. By a perspectival shift I mean a shift in McTaggart's tensed A-series {past, present, future}. By a metrical shift I mean a change in duration along a timeline ordered by the B-series {earlier than, later than}. In standard future discounting experiments, subjects are offered a change in the *metrical relationships between goods* but not the *temporal perspective*. Subjects are typically given a choice between a future-smaller-sooner reward (say, $20 next week) and a future-larger-later reward (say, $100 in two weeks). Both are delivered in the future, the same perspective. Until recently, no one has studied the metrical aspects of past discounting. When discussing past discounting, Prior [1959] and Parfit [1984] mostly concentrate on only the tensed perspectival shift. Once a present-headache has become a past-headache there is assumed to be a steep drop-off in my care for it. For example, I'd pay $20 for my present-headache to be relieved, but nothing, or perhaps a few cents, for my headache in 1978 to be relieved. The metrical aspect between past rewards is ignored.

Many of the most discussed impurities attach to the metrical features of time. In my paper I quote Ziff [1990] who says that time has a character, and that character can be pleasant, unpleasant or neutral. The metrical aspects of time have characters and not so obviously the perspectival aspects, which are often regarded as indexical in nature. An event being *now* doesn't alter its colour, loudness, and so on. An event being *long*, by contrast, can strongly affect all manner of qualities. In the standard discounting experiment you're asked if you want the reward in (say) one week or two weeks. Those two durations have different characters, characters deeply entangled with all aspects of your

life. The metrical aspects carry a lot of impurity. But if we only focus on the perspectival shift as opposed to the metrical shifts in past discounting, as Prior does, perhaps past discounting is more pure. And perhaps Miller is regaining a lot of the variation found in future discounting by now including metrical aspects in past discounting too.

However, perhaps I missed the mother of all impurities! As we navigate through life, we employ the tensed model I call *manifest time* [Callender 2017]. In it the past is viewed as fixed and the future open. Not only that, but psychologists have also shown that many other features are associated with these perspectives; for example, the future is felt to be brighter and lighter than the past [Helzer et al 2012]! According to manifest time the temporal perspectives are not simply indexical. They too have characters, e.g., the future is malleable. And unlike uncertainty, anticipation, and so on, manifest time colours *all* events.

I wonder if Miller empirically can untangle the signal from manifest time from the other 'metrical' impurities? Take two equally uncertain rewards, one in the past (to later be revealed) and one in the future. Does the 'openness' of the future amplify the uncertainty of the future, uncertain outcome?

Finally, normatively, here we meet another challenge of defining the pure/impure division. For someone who thinks manifest time gets time right, these impurities are in fact pure. That's one way of looking at Prior's [1959] famous 'thank goodness that's over' argument. Past discounting for him is picking up on a strong pure signal from time. For someone who doesn't think manifest time corresponds to what time fundamentally is, manifest time irrationally modulates our preferences. I try to rescue the rationality of some types of future discounting by saying that what is thought to be pure is really impure. Prior, by contrast, tries to rescue past discounting by saying the preference is pure and not impure.

Tim Smartt

I suggest that a principle of non-arbitrariness might be used to motivate the Invariance condition. Invariance plus Consistency will imply Stationarity, which together with a few more assumptions will imply that one should discount exponentially. I then criticize the principle of non-arbitrariness, hoping to weaken the motivation for Invariance. In his first point, Smartt [2022] wants to defend this principle. The principle holds that arbitrary features of a good shouldn't matter to one's preferences about that good, and as it relates to time, the idea is that temporal position, all else being equal, is arbitrary.

I have a preference to have a donut now instead of waiting until next week. Is this based on an arbitrary fondness for today? Only if all else is equal. But it isn't. I'm having a dental cleaning in a few days. I don't want to ruin that feeling of clean teeth with all that sugar. Best to have the donut now and let the dentist clean the teeth after they have been sugar-coated. To have normative bite the ceteris paribus clause must be discharged, the time preference must be pure. I point out that all else is never equal. I don't think that anyone has ever had a truly pure time preference.

Smartt defends the principle by insisting that some impure temporal preferences can be arbitrary. Suppose I want the donut today rather than next week because waiting makes me anxious. That's impure. But it could still be irrational, for perhaps I *assign the wrong weight* to the anxiety. It is miscalibrated, and in some sense, arbitrary.

As Smartt expects, I'm not moved by this. I see why he makes this move. The original target article sometimes gives the impression that one could wave impurity around like a magic wand and erase irrationality. But that's not what I think. I agree that some impure temporal preferences make sense and some do not. As mentioned in the introduction, we need a new normative standard to help us with this division.

Arbitrariness does not help. Smartt's rescue of the principal only works by equivocating on 'arbitrary'. Using the original meanings, now the principle of non-arbitrariness rules out arbitrary features *and* non-arbitrary features (when incorrectly calibrated) mattering to our rational preferences. The two features are joined under the arbitrariness umbrella only by calling formerly non-arbitrary features 'arbitrary' because they are poorly calibrated. Calibration now carries an enormous weight and it has nothing to do with the original motivation or meaning of 'arbitrary'. In any case, let's keep our eye on the ball. The original non-arbitrariness principle motivated Invariance and helped us derive exponential discounting. Does the new one?

I'm more sympathetic to Smartt's second point and I think it leads in interesting directions. He compares the move I make to one John Broome [1991] contemplates when considering the rational requirement that our preferences be transitive. In the example (see Smartt's paper), Maurice seems to violate transitivity and therefore be irrational. However, we could also re-partition the preference space and describe Maurice's preferences as transitive. Broome is worried that transitivity could become an empty requirement if we can always re-describe the case so that the normative condition is met. Smartt interprets me as re-describing pure temporal preferences that violate the normative standard as impure temporal preferences that do not violate the standard. He then points out that just as Broome doesn't go from the modest claim —that it's possible to re-describe—to the stronger claim—that this impugns the normative standard—one need not go from my modest claim to my stronger one. In fact, he says that I need to justify crossing this additional gap.

I thank Smartt for this point because the analogy is useful. I can say that I didn't think of myself as deploying a re-description strategy. Unbeknownst to me, I was implicitly adopting Dreier's [1996] response to Broome. Dreier holds that there is a fact of the matter about what preferences an agent has. I do too. I do not think we have something like underdetermination of theory by data here. We do not have two equally good descriptions of the agent, one with pure and one with impure preferences.

Smartt is right to say that the difficulty of determining purity or impurity doesn't by itself impugn the normative standard. But that's not exactly what I'm saying. I think I wasn't entirely clear on this matter. I'm claiming that the actual preferences are impure. EDU is a standard only for pure temporal preferences, so EDU is not the right standard for our actual preferences. What is the right standard for our actual impure temporal preferences is then an open matter. My introduction describes some options. Thanks to all these commentaries I'm now much clearer on these options than I was when I wrote the article. And thanks to Smartt I've been pushed to see how this debate maps onto other controversies about rationality and failures of people to conform with expected utility theory.

In fact, one can conceive of an objection in the neighbourhood of what Smartt describes. Smartt is pointing to the gap between the normative and descriptive. No amount of descriptive non-adherence need undermine the normative standard. In ethics, for instance, it's hard to be good, but that need not undermine one's conception

of the good. Or should it? Suppose the definition of what's good is so hard, unattainable and even unfathomable that it is irrelevant to ordinary moral judgements. Then it might impugn the standard. In epistemology this is the motivation behind the school known as 'bounded rationality' (e.g., Simon [1957], Cherniak [1986], Gigerenzer [1996], Harman [1986]): the normative standard, e.g., probabilistic consistency, is said to be so demanding as to be irrelevant. One could entertain a similar argument in this case against EDU. For my practical aim, I don't have to make a stand on this issue; but thanks to Smartt's comments I see new ways of developing my argument.

Preston Greene

Greene [2022] is unhappy with my lumping together philosophical temporal neutralists and social scientists who advocate for exponential discounting. The philosophers don't permit temporal discounting *at all* and yet our generalized 'social scientists' (for narrative convenience) permit *any* time discounting (so long as it's exponential). The former want $\rho = 0$ whereas the latter allow $\rho \neq 0$ so long as $e^{-\rho(t - \tau)}$. They're enemies, he says, not bedfellows!

Greene is of course correct that classifying them together does indeed obscure an important difference, namely, whether non-zero discount rates are permissible. I don't see that much hangs on this because this philosophical dispute isn't what my [Callender 2022a] paper is about. My paper is about the 'exponential vs hyperbolic' debate, a truly massive debate with significant policy decisions hanging in the balance. For the purposes of the paper I bracketed the question of whether rationality dictated that $\rho = 0$. For this debate it doesn't matter what value ρ has. And if I assumed that $\rho = 0$ I can't even state the debate.

As I was searching for inspiration for the condition Invariance, I did find inspiration in the temporal neutralists such as Sigdwick, Rawls and others. Invariance says that one's preferences should be independent of calendar time. I don't think it's a stretch to say that temporal neutralism suggests that. Greene says that their arguments do not support exponential discounting. That is probably right, although I'm only concerned with Invariance in this context. Given the unimpeachable fact that not discounting is exponential discounting ($e^0 = 1$), I'm not sure I see an issue. That the temporal neutralists' arguments sanction a stronger conclusion doesn't mean that they don't also sanction the logically weaker one.

I can offer Greene a condition that would get him what he wants. Call it *Strict-Temporal Neutrality*. Using the terminological conventions from the paper, it states

Strict-TN $(\mathbf{x},t) \sim \tau\ (\mathbf{y},t) \Leftrightarrow (\mathbf{x},t') \sim \tau'\ (\mathbf{y},t')$

This condition says that if you're indifferent now ($t = \tau$) between $\boldsymbol{x}$ and $\boldsymbol{y}$ then you should be indifferent at any other evaluation point ($t = \tau'$) too. The temporal relationship between $\boldsymbol{x}$ and $\boldsymbol{y}$ doesn't matter. Someone who satisfies Strict-Temporal Neutrality doesn't care about when the goods are delivered in calendar time because ($\Delta_2 - \Delta_1$) doesn't matter, and this preference is stable across time because it doesn't matter when τ is either. Notice, interestingly, that the difference between Invariance and Consistency becomes trivial when Δ drops out. So Strict TN can motivate Stationarity *and* Exponential discounting; more than that, it will imply $\rho = 0$.

Unlike many of the other commentators, Greene wants to defend the pure vs impure distinction by appealing to *mixed* pure-and-impure preferences. His defence

is really challenging and interesting, something I've been mulling over since I received his reply. He entertains what we might regard as a kind of vectorial combination model of the *grounds of preferences*. I prefer the donut today rather than next week. He agrees that there are many impurities here: I have to factor in the uncertainty of receiving the donut, the psychological anguish associated with waiting, and so on. Maybe no preference is ever purely pure. But look at the grounds for the preference, the reasons one has for it. Imagine these grounds as little 'arrows' supporting the preference, some pure, some impure. Perhaps temporal nearness is never the sole ground for an actual preference, he suggests, but if it is a ground it is still irrational.

This move is distinct from but a little bit like Smartt's. My preference to have the donut today is impure. But Smartt might say that if the grounds for it include pure temporal nearness, then it is miscalibrated.

I struggle with how best to respond. In a fuller discussion, I would want to hear more about what a ground of a preference is. And for my practical goals, I'm curious whether grounds of preferences are something social scientists can ever determine. In the case mentioned of high BMI being linked with hyperbolic discounting, can we ever find out whether temporal nearness is a ground and cause of unhealthy behaviour? If not then it won't be the sort of thing upon which we can base policy. But more than all this, I'm not sure grounds for preferences work in this kind of vector compositional way. And I worry that the problem of distinguishing pure from impure just reappears at the level of grounding components.

I wish I could put my point more carefully, but let me go through an example so readers can get a sense of what I mean. I don't want to go to the beach today because it's *too far*. Is this preference irrational? Is it partly grounded on *pure spatial distance*? Well, I don't wish to go to the beach due to the traffic and parking challenges and an appointment later; otherwise, I'd go in a heartbeat. Put me in a Star Trek transporter and I'd go, and in fact then I wouldn't care if it was in San Diego or really far away. Spatial distance matters, but only in how it combines with all the other contextual variables (transportation, parking, etc).

Greene would answer: yes, but is (say) 10% of the grounds of my wanting to not go to the beach due to it being *purely* too far? This is where I struggle. On the one hand, I see what he means and also its appeal, but on the other hand I honestly don't know how to answer. Pure spatial and temporal preferences are over very exotic states of the world. Presumably so are grounds for preferences. We individuate events in space and time via all the matter and interactions associated with different events. Now we're being asked: hold all that fixed —*everything you use to identify spatial and temporal events*— do you still care about where or when some good is delivered? It is similar to being asked if what underlies your preference is a fondness for one labelling of spatial or temporal coordinates over another. I suppose I want to grant that, logically speaking, it's possible that someone have grounds for a preference like this. Yet the preference or grounds of preference is so exotic I can't see it being relevant to social science or policy.

To be clear, I'm not saying that impurities always make a preference rational. I'm saying that *all* temporal preferences are impure, some sets rational, some not. In the example of wanting to go to the fish and chip shop for food poisoning, I agree with Greene that adding the impurity 'because it's cheap' doesn't make the irrational preference rational. Yet adding 'because I'm a sadomasochist' might!

Greene's third point is an alternative spin on the history of discounting in economics. I agree with most of it, so I'll use my space to answer a question he poses for

me in footnote 9. It's also a nice way to draw the discussion to a close. Greene asks: am I temporal neutralist or not?

In Callender [2022b], I distinguish between tensed and tenseless versions of temporal neutralism. I am tempted by the tensed version. Imagine a lifespan drawn across a landscape, from birth to death. You can distribute goods across this lifespan. The best arrangement depends on the way the world is. The landscape has hills and valleys, making earlier-for-later sacrifices make sense here, later-for-earlier sacrifices make sense there. The hills and valleys are the impurities in calendar time. There is no escape from the hills and valleys. One also imposes a temporal perspective on this worldline, dividing it up into past, present and future. The division hangs on where you are on your lifeline. Imagine it as a kind of red 'you are here' dot moving up the lifeline. Tensed temporal neutralism states that your choice of when goods should be delivered to you should hang on the mountains and valleys and not where the red dot is.

Spelled out more clearly (see Callender [2022b; 2023] for details), I think this is an interesting and plausible version of temporal neutrality. Intuitively, it says that it's okay to let birthdays, age, personal identity, uncertainty, and more matter in how you distribute goods throughout a lifetime, but it isn't okay to let when the now is do so. The main problem with this view is that the little red dot—you—changes as you traverse the landscape. You are not apart from the world. So, this version of temporal neutralism is only good for as long as we can approximate you and your preferences as stable.

Disclosure Statement

No potential conflict of interest was reported by the author.

References

Becker, G.S., and K.M. Murphy 1988. A Theory of Rational Addiction, *Journal of Political Economy* 96: 675–700.

Bleichrodt, H., and A. Gafni 1995. Time Preference, the Discounted Utility Model, and Health, *Journal of Health Economics* 15: 49-66.

Broome, John 1991. *Weighing Goods*, Cambridge, MA: Wiley-Blackwell.

Callender, C. 2017. *What Makes Time Special?* Oxford: Oxford University Press.

Callender, C. 2022a. The Normative Standard for Future Discounting, *Australasian Philosophical Review* 5/3: 227–53.

Callender, C. 2022b. Is Discounting for Tense Rational?, in *Temporal Asymmetries in Philosophy and Psychology*, eds Christoph Hoerl, Teresa McCormack and Alison Fernandes, Oxford University Press: 114–48.

Callender, C. 2023 (forthcoming). A New Derivation of Exponential Discounted Utility, *Philosophy of Science*.

Cherniak, C. 1986. *Minimal Rationality*, Cambridge, MA: MIT Press.

Dreier, James 1996. Rational Preference: Decision Theory as a Theory of Practical Rationality, *Theory and Decision* 40: 249–76.

Foley, R. 1993. Working Without a Net: A Study of Egocentric Rationality, NY: Oxford University Press.

Gigerenzer, G. 1996. On Narrow Norms and Vague Heuristics: A Reply to Kahneman and Tversky, *Psychological Review* 103/3: 592–6.

Grayot, J.D. 2020. Dual Process Theories in Behavioral Economics and Neuroeconomics: A Critical Review, *Review of Philosophy and Psychology* 11: 105–36.

Greene, Preston 2022. 'Pure' Time Preferences Are Irrelevant to the Debate over Time Bias: A Plea for Zero Time Discounting as the Normative Standard, *Australasian Philosophical Review* 5/3: 254–65.

Grüne-Yanoff, Till 2022. In Defence of Intertemporal Consistency. A Discussion of Craig Callender's 'The Normative Standard for Future Discounting', *Australasian Philosophical Review* 5/3: 266–76.
Harman, G. 1986. *Change in View: Principles of Reasoning*, Cambridge, MA: MIT Press.
Harvey, C.M. 1994. The Reasonableness of Non-Constant Discounting, *Journal of Public Economics* 53: 31–51.
Helzer, E.G., K.C. Hanko, and T. Gilovich 2012. Common Associations to the Past and Future and their Role in the Past/Future Attributional Asymmetry. Unpublished manuscript.
Ikeda, S. 2016. *The Economics of Self-Destructive Choices*. Springer Japan.
Janssens, W., B. Kramer, and L. Swart. 2017. Be Patient When Measuring Hyperbolic Discounting: Stationarity, Time Consistency and Time Invariance in a Field Experiment, *Journal of Development Economics* 126: 77–90.
McTaggart, J.M.E. 1908. The Unreality of Time, *Mind* 17/68, 457–74.
Metcalfe, J., and W. Mischel. 1999. A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower, *Psychological Review* 106/1: 3.
Miller, Kristie, James Norton, and Andrew Latham 2022. Pure and Impure Time Preferences, *Australasian Philosophical Review* 5/3: 277–83.
Parfit, Derek 1984. *Reasons and Persons*, Oxford: Oxford University Press.
Prior, A.N. 1959. Thank Goodness That's Over, *Philosophy* 34: 12–17.
Richards, T.J., S.F. Hamilton, and G.M. Pofahl 2010. Obesity and Hyperbolic Discounting: An Experimental Analysis, Agricultural and Applied Economics Association Annual Meeting, 25–27 July 2010, Denver, Colorado 61186.
Robertson, Ian 2022. High Time for a Change? A Response to Callender on Rationality and Preferences, *Australasian Philosophical Review* 5/3: 296–301.
Simon, H. A. 1957. *Models of Man: Social and Rational*, New York, NY: Wiley.
Smartt, Tim 2022. Arbitrariness Arguments against Temporal Discounting, *Australasian Philosophical Review* 5/3: 302–8.
Steele, Katie 2022. Why Time Discounting Should Be Exponential: A Reply to Callender, *Australasian Philosophical Review* 5/3: 284–95.
Strotz, R.H., 1955. Myopia and Inconsistency in Dynamic Utility Maximization, *Review of Economic Studies* 23: 165–80.
Ziff, P. 1990. Time Preference. *Dialectica* 44: 43–54.