Australasian Philosophical Review

Volume 5 Issue 3 September 2021

Routledge
Taylor & Francis Group

ISSN: 2474-0500 (Print) 2474-0519 (Online) Journal homepage: www.tandfonline.com/journals/rapr20

In Defence of Intertemporal Consistency. A Discussion of Craig Callender's 'The Normative Standard for Future Discounting'

Till Grüne-Yanoff

To cite this article: Till Grüne-Yanoff (2021) In Defence of Intertemporal Consistency. A Discussion of Craig Callender's 'The Normative Standard for Future Discounting', Australasian Philosophical Review, 5:3, 266-276, DOI: 10.1080/24740500.2021.2112122

To link to this article: https://doi.org/10.1080/24740500.2021.2112122

Published online: 09 Feb 2023.

Submit your article to this journal

Article views: 59

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=rapr20

AUSTRALASIAN PHILOSOPHICAL REVIEW
2021, VOL. 5, NO. 3, 266–276
https://doi.org/10.1080/24740500.2021.2112122

Routledge
Taylor & Francis Group

Check for updates

In Defence of Intertemporal Consistency. A Discussion of Craig Callender's 'The Normative Standard for Future Discounting'

Till Grüne-Yanoff

Royal Institute of Technology in Stockholm

ABSTRACT

While broadly in agreement with the conclusion that the exponentially discounted utility model (EDU) is not a universally valid rationality standard, I want to defend some intertemporal rationality criteria related to EDU, which Craig Callender might not share. My commentary explores the tension between these intuitions and Callender's arguments. In the first place, I show that many of the concerns that he raises are in fact compatible with intertemporal consistency (and sometimes even with EDU). Secondly, I rebut those arguments that directly challenge consistency. I conclude that while we disagree on the rationality of consistency, we can nevertheless agree on its non-universality and its limited justification of policy interventions.

ARTICLE HISTORY Received 2 June 2020; Accepted 8 June 2020

KEYWORDS intertemporal utility; exponential discounting; rationality; consistency; invariance; behavioral policy

1. Introduction

In his lead article, Craig Callender [2022] offers a lucid analysis of an important but underdeveloped domain of decision theory: the validity of EDU as a normative standard of intertemporal rationality. I am broadly in agreement with his conclusion that EDU is not a universally valid standard, but instead is dependent on contextual factors as yet unspecified; and that many of the current policy recommendations derived from this purported normative standard are questionable. Yet at the same time, I have robust intuitions about at least some requirements of intertemporal rationality—some of which Craig may or may not share. My commentary explores the tension between these intuitions and Callender's arguments. In the first place, I show that many of the concerns that he raises are in fact compatible with intertemporal consistency (and sometimes even with EDU). Secondly, I rebut those arguments that directly challenge consistency. I conclude that while we disagree on the rationality of consistency, we can nevertheless agree on its non-universality and its limited justification of policy interventions.

CONTACT Till Grüne-Yanoff gryne@kth.se

© 2022 Australasian Association of Philosophy

2. The Rationality 0f Avoiding Preference Reversals

Preference is a binary relation between two relata. They typically represent an agent's ranking of these relata in terms of her all-things-considered subjective evaluations. Preferences are temporal in at least two ways. First, many preference relata are realized at a specific point in time—for example dining at different restaurants *tonight*. Second, the agent forms the evaluation expressed in the ranking at a specific time—for example when comparing those restaurants *now*. Because agents' preferences are temporal, they can change. Many different factors cause preference change, but I want to focus on a particular one: the effect that a change in delay between evaluation and relata realization has on preferences. Such effects have been documented in the psychological literature originally giving rise to the hyperbolic discounting function (for example, Kirby and Herrnstein [1995], Ainslie [2001]). In want of a better name, let me call them *delay-dependent reversals*. I am convinced that delay-dependent reversals are irrational, in the sense that we have good reasons to either prevent them from occurring or from influencing our decisions, for the following reasons.

First, in my personal experience, delay-dependent reversals are disempowering. I form intentions and plans, only to later feel that I do not want to realize them—without any discernible external influence (no new information, no new experience, no new insight) making it so. Once the time for realizing them is passed, however, I come to regret not having realized them. Of course, such reversals often are not surprising: I often anticipate that they will occur, and that I will regret having had them later on. Many people seem to share this experience. George Ainslie [2001: 40] vividly describes the agony of the addict here:

> Often I'll choose in the opposite direction when I'm close and when I'm distant, which means I'll regularly do things at one time and undo them at another. Obviously if what I do in a particular situation regularly gets undone, I'll learn to stop doing it in the first place—but not out of agreement with the later self that undoes it, only out of realism.

Here, the individual is at war with herself: conflicting intrapersonal interests, controlling behaviour at different times, strategically interact to maximize their share of reward. Earlier interests might learn to deprive later interests of their controlling power—through commitment devices, for example, through distraction, or through establishing internally stabilized personal rules. But the later interests are just as capable in outwitting these moves of the earlier selves, overcoming obstacles and distractions, and destabilizing personal rules ('just this once, it's such a special occasion!'). The deep sense of frustration, the damage to self-estimate, the sense that one lacks the power to determine the course of one's life as a whole, instead of it being merely the outcome of a contingent sequence of skirmishes, I believe, is a strong prima facie reason to prevent delay-dependent reversals.

Second, beyond the subjective experience of disempowerment, delay-dependent reversals also prevent the realization of one's plans. Admittedly, not all long-term plans are good or desirable. But avoiding preference reversal is a necessary condition for realizing *any* long-term plan.[1] Thus, while successful planning doesn't make projects worthwhile, many projects that are worthwhile could not be realized without

[1] Of course, some of one's preferences can reverse without undermining a given plan. But the realization of any plan requires that at least some preferences don't reverse, or at least that this reversal has no effect on one's behaviour.

successful planning. The ability to realize such long-term plans is often seen as an important part of rationality [Bratman 1983, McClennen 1990]. Avoiding preference reversal, because it is a necessary condition for the possibility of realizing long-term plans, becomes part of this rationality condition—thus making them rational in their own right.

Third, both coordination and cooperation between persons depends on each forming beliefs about others' motivations. To coordinate actions without verbal communication requires that agents predict each other's preferences. To achieve stable cooperation (with or without communication) requires that agents assess their opponents' incentives to defect, which in turn requires them to predict each other's preferences. Intertemporal stability and the absence of preference reversals increases this predictability and thus facilitates coordination and cooperation. To the extent that that coordination or cooperation is desired, avoiding preference reversals is rational.

Fourth, finally, there is widespread agreement that moral accountability presupposes personal identity. One can be responsible only for one's own actions, and what count as 'one's own' requires reference to personal identity (e.g., Parfit [1984]). What such identity consists in is controversial, and intertemporal stability of preference is not the main candidate in any of the extant accounts. However, it might play important roles in shoring up various prominent accounts. For example, to render Lockean psychological continuity more plausible, some have argued one should incorporate more psychological features than just memory into the identity-preserving relation—'present-future relations such as intentions fulfilled in action, relations that persist across time such as beliefs, goals, and desires' [Shoemaker 2019]. Alternatively, narrative accounts require for personhood that experiences and psychological characteristic be woven together through a story that the subject tells, giving them coherence they otherwise wouldn't have had (e.g., Schechtman [1996]). Either of these accounts seem to require the absence of delay-dependent preference reversals, and so, to the extent that these forms of personal identity are desirable, it is rational to avoid them.

Each of these arguments depends on further assumptions that of course can be questioned, but overall they provide what I consider a pretty robust intuition that it is rational to avoid delay-dependent preference. It seems that Callender [2022: 245] shares at least some of these intuitions, as when he says 'after all, one wants to say that being able to make and stick to a plan is a sign of rationality'. Yet at the same time, some of his criticisms contradict these intuitions.

3. Callender's Rejection of EDU

Callender rejects the claim that EDU is a necessary condition of rationality. In this section, I deal with those of his arguments that I believe are compatible with my intuition that delay-dependent preference reversals are irrational.

EDU is a property of the discount function within the discounted utility model. The extant representation theorems identify necessary and sufficient conditions of preferences to be representable by such a utility model, and for such a utility model to be the unique (up to positive linear transformations) representation of such preferences. Furthermore, there are necessary and sufficient conditions on preferences so that the utility representation of these preferences satisfies EDU. Between

my intuition and EDU there might be plenty of space to accommodate Callender's criticism.

The stationarity axiom is the crucial condition on preferences for EDU. Stationarity, however, is hard to evaluate normatively—it just runs together too many disparate features to compare easily to common intuitions. Fortunately, as becomes clear in Callender's lucid analysis, the two necessary and jointly sufficient conditions for stationarity are consistency and invariance. Roughly put, consistency prohibits preference reversals due to changes in the temporal distance between agent and realization of preference relata—that is, delay-dependent reversals. Invariance prohibits preference reversals due to absolute changes in calendar time.

All my intuitions from section 2 are thus about consistency, not invariance. Thus, what I need to defend against Callender is consistency, not invariance. In this section, I do so by pointing out which of his criticisms do not concern consistency. In the next section, I rebut criticisms directly aimed at consistency.

Callender correctly notes that violations of EDU are often incorrectly associated with some other purportedly undesirable feature, and that the normative reprobation wrongly arises from this association. Mischel's Marshmallow test is perhaps the most prominent example, in particular as there is now a whole literature how 'failing' this test is predictive of 'failing' in real-life tasks requiring patience, endurance, resolve, etc. Callender is of course right that the Marshmallow test does not test EDU violations, and thus its association with 'real-life failings' has no relevance for the evaluation of EDU (and of consistency in particular). But I disagree with his claim that this association 'drives much of the moralizing' [ibid.: 241] about EDU violation. None of the above reasons for my intuition rest on any such association; hence even if claims about such association are dropped entirely, the normative questions prompted by EDU violations remain.

Callender then turns the tables and wonders if the direction of causality isn't the other way around: 'Which way this causal arrow goes between time preference and personal and social outcomes is still contested. It is a tension intimately connected to the choice of normative standard for time preference' [ibid.: 230]. The argument here remains somewhat vague. It might indeed turn out to be the case that poverty or other adverse conditions cause EDU violations. But this direction of causality is compatible both with EDU violations being irrational and with them being rational. Poverty-induced EDU violations might be as irrational as stress-induced miscalculation or alcohol-induced rage-attacks [Mani et al. 2013]. Or, in conditions of poverty, it might be sufficiently prudent to overvalue the present that EDU violations should be considered rational. Evidence for causal direction thus does not provide basis for the normative evaluation of EDU violations. Rather, a separate argument must be furnished that under such adverse conditions, EDU violations are indeed rational. That Callender does not offer.

My intuitions for judging consistency rational do however appeal to some of its consequences—the success of plans, of coordination/cooperation and of moral accountability. Because these are desirable, I argued, it is irrational to violate consistency. Callender challenges this desirability assumption: not all plans are desirable—successfully realizing those might plot 'paths to sure ruin' [Callender 2022: 241]—and deferring gratification is not universally desirable either [ibid.: 241–2]. I am not convinced by these objections. Any instrumental reasoning can be used for malicious ends; yet absent a consensus on the evaluation of ends, it has been very fruitful to

evaluate instrumental reasoning in isolation from the ends for which it might be employed. Thus, dynamic consistency may be considered rational even though this rationality also will help malicious plans to be realized.

Regarding gratification deferment, I agree with Callender: one can definitely overdo it (and in fact, in my opinion, many *do* overdo it). But this implies nothing about the rationality of consistency, for two reasons. First, although consistency requires absence of delay-dependent preference reversal, it does not say that the earlier, long-term preference is the rational one that mustn't be changed. To the contrary, it might be rational to adjust one's earlier, long-term preference in the light of the expectation that one's later, short-term preference will be different and more relevant for oneself. Consistency, properly understood, is neutral with regards to gratification deferment.

Second, excessive deferral is not caused by consistency. In cases where agents are consistent, their very low discount rates are to be blamed. The anorexic values future slimness so much as to accept adverse health effects now. The miser values the opulent savings account so much as to accept adverse social effects now. Many people judge such behaviour because they think the trade-off is bad. To 'cure' the anorexic or the miser, they want them to value health and sociality more, and slimness and savings less. But changing the trade-off is a matter of changing the *rate* of discounting (compatible with EDU), not its form. In fact, it seems to me that some of those who excessively defer are violating consistency: they often seemed to have long-term preferences to eat or spend more, yet when facing the concrete choice, they choose differently. Thus, neither miser nor anorexic would be helped by an EDU-violating discounting function; instead of softening their social or health strictures, such a function would only increase their misery by adding inconstancy, futility and regret to the mix.

Callender [ibid.: 242] also worries about the 'evolutionary story associated with the standard [EDU] model'. Apparently, he has various dual-process accounts in mind here. I agree that these accounts are bad. Methodologically, proponents have been unable to show that these processes indeed have stable clusters of properties (for example, fast/automatic/non-malleable vs. slow/deliberate/malleable). Conceptually, it remains a mystery who decides which processes get to decide at what time. But this is discussion is a red herring, anyway. How is an evolutionary story going to help us decide whether EDU is necessary for rationality or not? If intertemporal consistency were already selected for, we wouldn't need to have this discussion—not because adaptiveness implies rationality, but because normative requirements become uninteresting if they are habitually satisfied. It seems however that humans don't come with intertemporal consistency already built in. There *might* be an evolutionary story why that is—perhaps along the familiar 'adaptive for our ancestors, but not for us' lines. But there might not. It might just be a brute fact of our physiognomy that we perceive delay not as absolute amount, but as a ratio of the change in the prior amount—as psychophysicists have detected in many other senses [Ainslie 2001: 35n16]. Humans might need to correct these perceptions, just as children need to learn to adjust size perception at a distance. Whether this is a valid normative requirement depends on the consequences it has for relevant purposes, not whether there is an evolutionary story or whether it was adaptive in some previous context.

To conclude, in this section I responded to Callender's criticisms that could be construed as criticizing consistency, but at closer analysis turn out not to. Now let's turn to those that genuinely do challenge my intuitions.

4. Callender's Criticism of Consistency

Most of my intuitions for the rationality of intertemporal consistency draw on the possibilities it generates: the possibility of empowerment, successful planning, cooperation, and moral accountability. Consistency is a necessary condition for these desirable possibilities, and therefore it is rational to maintain it.

Callender challenges such arguments by attacking this necessity assumption (specifically for successful planning, but it applies *mutatis mutandis* to the other two arguments as well). He does so in three ways. First, he argues, as the domain of the utility function is unrestricted, its 'hyperbolic parts' might cover time periods that the agent will not actually experience, so that the changes in her temporal relation to the relata cannot affect the preferences: 'Absent other information, a life without preference reversals can be modelled with an exponential or hyperbolic function' [Callender 2022: 242].

Second, he argues that even if the 'hyperbolic parts' of one's utility function cover actual time periods of an agent's life, the preferences thus represented might never be realised, thus never leading to any preference reversal. Humans have many preferences that they never act on, due to a lack of opportunity. If the preferences that would lead, if realized, to temporal inconsistency, are systematically not acted on, then the normative pressure to avoid such preferences should also ease up.

Third, Callender develops this thought further to counter exploitability claims. Such arguments point to a possibility: that people with inconsistent preferences *might* be exploited (typically financially) by clever adversaries. But what if there aren't such adversaries, or people who have inconsistent preferences behave in ways that prevents such exploitations? Then it is only the potential exploitation that supports the rationality of intemporal consistency, and that is not very strong. These arguments are well-rehearsed with respect to the rationality conditions on beliefs. Callender concludes 'the connection between exploitation and irrationality is very tenuous. It's tenuous in the case of credences and at least as tenuous in the case of preferences' [ibid.: 244].

These are serious objections, but I believe I can counter each. First, it is helpful to distinguish between the abstract normative concept of rationality and the practical prescriptions that may or may not flow from it. Surely, I sometimes form irrational beliefs —for example, that are not consistent with each other, do not satisfy the probability axioms, or are incorrectly inferred from the evidence. Yet in many cases, it isn't worth the effort to clear out these irrationalities; they don't do any identifiable harm, and such clearing-out might produce unwanted side-effects. Should one therefore say that those cases not worth the effort shouldn't count as irrational, either? No: they remain normatively deficient, but instead become prescriptively inert. Of course, rationality remains tethered at least partly to prudential relations that are justified in pragmatic scenarios. But while rationality conditions thus receive their justifications through such pragmatic contexts, they can be extended beyond them. The advantage is a cleaner concept of rationality, which need not carry the substantial baggage of context dependence. The disadvantage is that the normative rationality judgment does not automatically imply a prescription or the justification of intervention.

This distinction can be applied to all three of Callender's arguments. The irrationality of temporal inconsistency is justified with reference to actual cases of psychological instability, failed planning cooperation and moral accountability. The sufficiently robust intuitions for those actual cases justify the extension to possible cases: it is

irrational to be temporally inconsistent also in the evaluation of counterfactual scenarios, of situation one can prevent encountering or that have no negative practical consequences. However, this implies that any prescriptive and practical conclusions from such normative judgments require further justification.

Second, I disagree that the connection between exploitation and irrationality is more tenuous in the case of preferences than in the well-rehearsed belief cases. Callender explicitly bases his objection on the epistemically demanding assumptions of the exploitation arguments: 'she is only exploitable by someone who knows that she will change his mind and the price points at which she will. We typically don't encounter such people, so it might be rational to assume we won't meet any' [ibid.]. That might be true for interpersonal scenarios, common in the discussion of rational belief ('clever bookies'). But with respect to preferences, perhaps the more important scenarios are *intrapersonal*: conflicting interests that all vie for reward but that have temporally limited ability to steer behaviour.[2] Importantly, epistemic limitations do not apply to the intrapersonal scenarios in the same way as to interpersonal ones: housed within a person, and continuously interacting with each other, these interests are intimately familiar with the particularities and potential weaknesses of their adversaries, and they will exploit them. The epistemic argument thus applies *less* to preference than to belief cases, and therefore, it is not correct that the exploitability argument is at least as tenuous in the former as in the latter.

Third, Callender offers another argument against exploitability: that people might find ways not to act on those preferences that would generate preference reversals. Such strategies are indeed important ways to overcome preference reversals and their negative consequences. People learn to overcome procrastination, or binging, or failing to save, by foreclosing choice options or preventing themselves from choosing during those time periods in which preferences conflicting with their plans prevail. But the successful employment of such strategies is *not* an argument *against* the irrationality of preference reversals. To the contrary: what justifies the implementation of these strategies is precisely the judgment that the otherwise threatening preference reversal is irrational. The normative evaluation based on the potential exploitability justifies the steps that prevent this exploitation. This shows both why it is important to base rationality judgment on potential, not actual exploitability, and why this rationality judgment is an important precondition to successfully avoid exploitation.

I therefore conclude that consistency is a necessary condition for the desirable possibilities spelled out in section 2, and that Callender's challenges to this claim do not succeed. For these reasons, temporal consistency is a necessary condition of rationality.

Callender secondly challenges the notion of a *purely temporal* preference reversal. All my intuitions depended on my ability to distinguish delay-dependent preference reversals from reversals caused by other factors. As I pointed out in section 2, there needs to be nothing irrational about changing one's preferences due to learning, refinement, epiphany, forgetting or ageing, to give just some examples. Only those preference that reverse merely due to the temporal position of a prospect vis-a-vis the subject are considered irrational. But can these reversals be properly delineated?

[2] Accounts of multiple competing temporal selves have been developed in parallel in psychology (e.g., Ainslie [2001]) and economics (e.g., Laibson [1997]). Unlike dual-process accounts, these accounts explicitly model how temporally distributed selves strategically interact in order to achieve behavioural control.

Callender thinks not. People change their preferences continuously due to such extraneous factors, and they do so in time. Changing one's distance to the preference relata might thus always be intermixed with exposure to those factors, so that in fact people never experience preference reversal due to one or the other cause only. Such actual intermixtures, Callender [2022: 245] suggests, prevent the application of intertemporal consistency as a relevant rationality criterion:

> In the face of all this flux, is there anything we can insist on remaining invariant . . . What's clear is that we're not going to come through all of those moves with EDU intact as the normative standard. In each case we must retreat to something invariant, thus limiting the scope of EDU in problematic ways.

I can think of three ways how to limit the scope of EDU in response to this intermixture argument.

i. *actual integration:* all preference reversals are irrational. Rationality individuals must foresee influence of external factors and either adjust their earlier preferences or take effective steps to counteract the influence of these factors.
ii. *actual separation:* preference reversals are irrational only if caused by nothing but delay changes. If additional factors have influence, all bets are off.
iii. *theoretical separation:* preference reversals are irrational only if they are (partially) caused by temporal distance changes, *ceteris paribus*. This emerges from theoretical analysis.

I agree with Callender that (i) is far demanding, and (ii) would make the principle of intertemporal consistency inapplicable. But I see (iii) as a way forward, while Callender is more pessimistic. He focusses on a particular version of (iii) that makes all side conditions explicit in a set of maximally specific possibilities, and interprets consistency as holding only over these. But that, he claims, 'would make the theory way too ideal for economics. The ultimate/non-ultimate divide is untestable' [ibid.].

I disagree with Callender on how problematic such idealizing assumptions are. Idealizations in normative theory are different from idealizations in positive theory. Normative idealizations describe how the world ought to be, not how it is. If the normative theory only made claims that are true anyway, it would be irrelevant. Normative theory thus at least partly is *relevant because* it makes false claims about the world. Positive idealization, the typical form of idealization in scientific models, stands in conflict with the purpose of accurately representing the target. It be shown to be *relevant despite* making false claims. Therefore, that a normative theory is idealizing is not a prima facie problem for that theory, and the conditions for its relevance cannot be simply subsumed under the widely discussed relevance conditions for positive idealization (for methodological discussions of normative theory evaluation, see Guala [2000] and Mongin [2006]).

Admittedly, Callender challenges the *c.p.* conditions, not the normative claim itself - it is them that are 'way too ideal' and 'untestable'. They qualify the conditions under which the normative claim is valid. Normative claims that come with a lot of qualifications are less desirable, first because the qualifications reduce their relevance for the actual situations that humans find themselves in; and secondly, because the qualifications make it increasingly difficult to support these claims with normative intuitions.

While I agree that these are in principle good reasons to be sceptical about idealizing *c.p.* conditions in normative theories, I don't consider them particularly threatening in the case of intertemporal consistency. The first step is to acknowledge that it is not necessary to make the *c.p.* conditions explicit, as — in the maximally specific possibilities framework that Callender considers. Instead, economist and other social scientists often successfully employ vague *c.p.* clauses [Hausman 1992]. Although it is hard to positively enumerate all the possible factors that might affect one's preferences as time is passing, I find it much more plausible that one is able to identify periods of time where no such factors were active (as Justice Stewart said: 'I know it when I see it'). We might say that preference reversals are irrational, except when legitimate external factors drive the reversal—and let legitimacy as depend on the specific context. This allows for easy applicability to situations relevant to humans, and provides the necessary flexibility without giving up on a general rationality principle.

The second step is to see that the intuitions I describe in section 2 already involve some kind of theoretical separability. The imaginary situations to which these intuitions apply already abstract from any factors that could legitimately drive reversals. So, the *c.p.* conditions of the idealized normative theory actually help to link the normative claim to the intuitions, rather than hindering it. Considering these two reasons, I do not think that consistency is 'too ideal'. And I am confused by the claim that the *c.p.* conditions are not 'testable'. This cannot be meant in the sense of theory testing—as there is not empirical data that could falsify the normative claim. Perhaps Callender means by it becomes unclear when the normative claim is supported (for example, by other theory, or by intuition). But that I argued is not the case. Finally, it might mean that the applicability of the normative principle is vague. While I agree with that, I have argued that there are practical ways to deal with such vagueness, and that such ways are in fact widely practiced in the social sciences. I therefore conclude that Callender's arguments against consistency are not convincing.

5. Conclusion

I have argued against Callender's objections to intertemporal consistency. But consistency is not sufficient for EDU. It also requires invariance. I have nothing to say about that. Instead, I believe Callender is right in rejecting invariance. People's preferences change with calendar time, and in contrast to the inconsistency case, I see no way how one can successfully separate illegitimate from legitimate factors for such change. Thus, Callender is right to conclude that EDU is not a universal rationality principle.

However, while he dismisses EDU on a broad set of concerns, I have considerably narrowed the basis of criticism—having rejected his criticisms of consistency, and also shown that some of his criticisms do not apply to consistency or EDU at all. This in turn opens the possibility of using EDU as a representation of intertemporal rationality in those situations where invariance-related concerns can be excluded. Yes, this is a 'retreat to safer grounds' as Callender observes, but that ground is still plenty to stand on.

A last point on policy relevance. Callender [2022: 250] is dismissive of 'time preference therapy' and 'pills to fix unhealthy time preferences' as actionable consequences of accepting EDU. I agree—but for reasons of my own. Instead of rejecting all general principles of intertemporal rationality, I maintain the normative validity of consistency

under vague *c.p.* conditions. But this complicates justification of interventions aiming to change people's behaviour without their participation. Whether in a given case the vague *c.p.* conditions are satisfied—that is, whether legitimate reasons for preference change are absent—is difficult to determine without the participation of those whose behaviour is supposed to be changed. The epistemic conditions for successfully and legitimately restoring consistency 'from the outside' are rarely met. Yet *c.p.* consistency remains a rationality principle, which gives both individuals violating it and policy makers seeking to prevent such violations reason to act.

Given the epistemic complications, restoring and maintaining intertemporal consistency becomes first and foremost a personal achievement. People themselves must overcome inconsistency in order to achieve the benefits of consistency. Engineering choice architectures to prevent preference reversals (e.g., Thaler and Benartzi [2004]), massively affecting incentives, or coercing people might achieve specific benefits for specific reversal problems, but does not achieve the benefits that I pointed to—a sense of empowerment, completing one's own plans, coordination/cooperation and moral accountability. The policymaker's role in this nevertheless can be substantial: for example, by providing choice contexts conducive to such an achievement, or fostering people's competences. From my view on consistency, therefore, policy does not come in the form of pills or therapies, but in active engagement with the people to be helped, nurturing, fostering, *boosting* their abilities to get themselves out of an irrational predicament [Hertwig and Grüne-Yanoff 2017].

By defending my intuitions, I insist on general principles of intertemporal rationality. It's not just a matter of contextual dependence and subjective tastes, as Callender seems to suggest in drawing the analogy to siting/spatial preferences. Instead, it makes sense to treat time as an exogenous parameter of preferences in order to explain people's choices; and there are good reasons to judge certain interactions between time and preference as *generally* irrational, *ceteris paribus*.

Disclosure Statement

No potential conflict of interest was reported by the author.

ORCID

Till Grüne-Yanoff http://orcid.org/0000-0002-6727-8723

References

Ainslie, G. 2001. *Breakdown of Will*, Cambridge: Cambridge University Press.
Bratman, M. 1983. Taking Plans Seriously, *Social Theory and Practice* 9(2/3): 271–87.
Callender, C. 2022. The Normative Standard for Future Discounting, *The Australasian Philosophical Review* 5/3: 227–53.
Guala, F. 2000. The Logic of Normative Falsification: Rationality and Experiments in Decision Theory, *Journal of Economic Methodology* 7/1: 59–93.
Hausman, D.M. 1992. The Inexact and Separate Science of Economics, Cambridge University Press.
Hertwig, R., and T. Grüne-Yanoff 2017. Nudging and Boosting: Steering or Empowering Good Decisions, *Perspectives on Psychological Science* 12/6: 973–86.
Kirby, K.N., and R.J. Herrnstein, R. J. 1995. Preference Reversals Due to Myopic Discounting of Delayed Reward, *Psychological Science* 6/2: 83–9.

Laibson, D. 1997. Golden eggs and hyperbolic discounting, *The Quarterly Journal of Economics* 112/2: 443–78.
Mani, A., S. Mullainathan, E. Shafir, and J. Zhao 2013. Poverty Impedes Cognitive Function, *Science*, 341/6149: 976–80.
McClennen, E. 1990. *Rationality and Dynamic Choice*, Cambridge: Cambridge University Press.
Mongin, P. 2006. A Concept of Progress for Normative Economics, *Economics & Philosophy* 22/1: 19–54.
Parfit, D. 1984. *Reasons and Persons*, Oxford: Oxford University Press.
Schechtman, M. 1996. *The Constitution of Selves*, Ithaca, NY: Cornell University Press.
Shoemaker, D. 2019. Personal Identity and Ethics, *The Stanford Encyclopedia of Philosophy* https://plato.stanford.edu/archives/win2019/entries/identity-ethics.
Thaler, R.H., and S. Benartzi 2004. Save More Tomorrow™: Using Behavioral Economics to Increase Employee Saving, *Journal of Political Economy* 112/S1: S164–S187.