PRUDENCE, BENEFICENCE, AND TIME[1]

Ben Bradley

wbradley@syr.edu

When an agent's values will change in the future, whether as a result of the agent's present action or in some other way, is it prudent for the agent to sacrifice their present values for the ones they will come to have later, if this will make their life better on the whole?[2] In what follows I argue, contrary to several extant theories of prudence, that the answer is "yes".

Before I begin, let me put to the side an obvious reason to discount one's future values to some extent, namely uncertainty about the future: uncertainty concerning whether one will continue to exist, what one's future values will be, or whether those values will be satisfied. A complete theory of prudence needs to deal with such uncertainties. But that is a concern of formal epistemology, not value theory; I shall assume omniscience on such matters in order to focus on the axiological issue.

I. TIME-NEUTRAL PRUDENCE

[1] For helpful discussion of the issues in this paper, I thank Krister Bykvist, Jamie Dreier, Adriano Mannino, Sean McKeever, Richard Pettigrew, two anonymous referees, and audiences at the University at Buffalo, New Mexico State University, and the Arizona Workshop in Normative Ethics.

[2] For some influential discussions of these issues over the last fifty years see Strotz 1973, Bricker 1980, Ullmann-Margalit 2006, Bykvist 2006, and Paul 2014. Throughout this paper I use the term "values" as a generic term intended to include desires, preferences, utility functions, beliefs about one's good, and the like; I take no stand on the attitudes to which a subjectivist should appeal in their theory of welfare.

One very simple view about prudence states that it is prudent for one to do what results in the best life for one. As Joyce Trebilcot put it: "the principle of prudence… is that one ought to maximize the sum of intrinsic value or minimize the sum of intrinsic disvalue for oneself over one's lifetime as a whole" (Trebilcot 1974, 204).[3] Call this the *Lifetime Welfare View of Prudence*. According to the Lifetime Welfare View, the theory of prudence is a trivial extension of the theory of well-being; once we have the correct theory of well-being, we have done all the interesting work there is to be done concerning prudence. Here I'll combine the Lifetime Welfare View with some common assumptions about well-being that together give rise to allegedly problematic implications.

It is common in discussions of prudence to assume that there is such a thing as how well-off an individual is at a particular time. We may combine the Lifetime Welfare View with a view about how lifetime welfare is generated by temporal welfare: an individual's lifetime welfare is the sum, for all times in that individual's life, of their welfare levels at those times. Call this the *Summative View of Lifetime Welfare*. The Summative View is sometimes rejected because of a commitment to sequencing or narrative effects, temporally extended goods, or goods that are difficult to locate in

[3] See Bykvist 2006, 269 for a similar statement in terms of outcomes rather than whole lives. Also see Brandt 1972, 681, and Parfit 1984, 3. Dietz seems to endorse this view implicitly in his criticism of Dorsey; he argues that Dorsey's view faces a tension in the way it cleaves total welfare from prudence (2023, 996). But simply wielding the view against the views I discuss in this paper would beg all of the interesting questions.

time.[4] I put these controversies to the side to streamline the discussion; assuming the Summative View helps us see exactly where time-relative views depart from time-neutral views.

It is also common in discussions of prudence to presuppose a subjectivist theory of well-being according to which an individual's well-being is determined by facts about what the individual values, in combination with facts about the extent to which the world fits those values. Let us take this to be a view about temporal well-being, as follows: an individual's welfare level at time t is determined by the extent to which the world matches the values the individual has at t, adjusted for the strength of those values. Call this the *Simple Subjectivist View of Temporal Welfare*.

The Simple Subjectivist View of Temporal Welfare leaves something undetermined: what is the relevant state of the world? Is it the state of the world at the time the individual has the values, or the state of the world throughout all of history? In this paper I will assume that it is the state of the world throughout all of history that is relevant, following e.g. Pettigrew (2019, 23, n3); I don't think the choice affects the issues I address here.

[4] Bykvist 2022 is a recent example. See Dorsey 2015 and Kauppinen 2015 for two attempts to account for narrative effects within a summative framework. Ben Bramble rejects the Summative View because he thinks there is no temporal welfare; I argue that Bramble's concerns can be accommodated by restating the views with no real loss of content (Bradley 2021). Gil Hersch argues that even if Bramble is right, temporal welfare is more useful in policy guidance (Hersch 2022).

A noteworthy feature of the combination of views just sketched is that when determining which life is best, all moments within the life count equally. According to this view, the prudential value of some future state of affairs is *time-neutral*; its prudential value does not change over time. For this reason let us call this combination of views (The Lifetime Welfare View of Prudence + The Summative View of Lifetime Welfare + The Simple Subjectivist View of Temporal Welfare) the *Simple Time-Neutral View*.

Another noteworthy feature of the Simple Time-Neutral View is that it does not matter on this view how it comes about that the world matches one's values – in particular, it does not matter whether one changes one's values to match the world or vice versa. What matters is the match. One acts prudently if one gets the world to match one's values, over one's whole life, to the greatest extent. If that requires changing one's values then so be it.

These two noteworthy features of the Simple Time-Neutral View strike many philosophers as wrong. According to some, present values count for more, or even for everything, in prudential deliberation; as Bernard Williams wrote, "The correct perspective on one's life is *from now*" (Williams 1981, 13). According to others, we do not have reason to change our values to ones that the world will better satisfy. As Phillip Bricker points out, that would require us to take a "complacency pill" that would cause us to be satisfied with whatever comes about (Bricker 1980, 398). And at least since Elster (1983), many of those writing about "adaptive preferences" have agreed with Bricker that this is irrational. I now turn to views that deviate from the Simple Time-Neutral View with respect to time-neutrality and the prudential reason to change one's values.

II. TIME-RELATIVE PRUDENCE

Parfit's well-known "Present-Aim Theory," which develops the subjectivist thought in a time-relative way rather than a time-neutral way, is an extremely time-relative view.[5] What is prudent, on this view, is what best satisfies the agent's values *as they are at the time of action* (Parfit 1984, 117). The values the agent has at other times are wholly irrelevant except insofar as they are themselves the objects of the agent's present values. So for example, if I do not presently value playing pickleball, there is no prudential reason at all for me now to bring about the state of affairs in which I play pickleball in the future while valuing it (unless I now value getting whatever I value in the future, or somesuch). So I have no prudential reason to do things to satisfy my future or past values, or to change my values to conform to ones that will be better

[5] Parfit does not present the Present-Aim Theory as a theory of prudence, but rather as a theory of rationality that is in competition with the Self-Interest Theory (aka the prudential theory) of rationality. Pettigrew also frames his discussion as about rationality rather than prudence. Dorsey does, however, take himself to be giving a theory of prudence. I am treating all of these views as views about prudence. If you think it is a conceptual truth that prudence solely concerns lifetime welfare, as suggested by the Trebilcot quotation above, you will think a time-relative view is not about prudence at all, but rationality. I refrain from making any such assumption in order to avoid begging any questions; but the arguments I present would apply just as well if we consider Parfit and Pettigrew's views to be views about rationality and not prudence.

satisfied. The Present-Aim Theory completely divorces prudence from lifetime well-being.[6]

The Present-Aim Theory seems to be presupposed by traditional approaches to decision theory. According to Yoaav Isaacs, defending simple expected utility theory, "It is easy to say what an agent should do in the face of potentially changing utilities—maximize expected utility relative to his current credences and current utilities… You needn't feel compelled to help your future self get what your future self will want" (Isaacs 1077-8).

There are less extreme ways to develop Williams's thought about the correct perspective on one's life. [7] Richard Pettigrew defends a view that charts a middle course between the Simple Time-Neutral View and the Present-Aim Theory. Pettigrew follows Parfit in holding that we should think of ourselves in the future and the past as something like distinct people with their own interests.[8] But the present self does not maintain absolute priority, as in the Present-Aim Theory. Just as when making moral decisions we ought to take the interests of other people into account, when making

[6] The "aprudentialist" theory of rationality described by Trebilcot (1974) is another example of this sort of view, but the aprudentialist is stipulated not to place any value on the future at all (Trebilcot 1974, 205), which is not part of the Present-Aim Theory.

[7] One such view is held by Paul Weirich; he claims that "a present aversion to the objects of future desires can justify discounting those future desires" (Weirich 1981, 31), and that "rationality is biased towards present desires" (37). I focus here on Pettigrew's more fully-developed view, which bears some similarity to Weirich's.

[8] See Bruckner (2004) for a slightly different way of thinking about the way prudence should account for past and future selves.

prudential decisions we ought to take the interests of our future and past selves into account and somehow aggregate the distinct values of these different temporal selves.[9] But we need not aggregate the way we do on the Simple Time-Neutral View. Just as when making moral decisions we may give extra weight to our own interests over those of others, when making prudential decisions we may give extra weight to our present interests over our future and past interests. Without the option of such a weighting, our decisions would be held hostage by our future values, which may be values with which we do not identify (Pettigrew 2019, 198). My present self, at the time I make a decision, carries special significance; to deny this would be an unacceptable relinquishing of autonomy or integrity, in much the same way that utilitarianism is alleged to require agents to relinquish what matters to them in order to satisfy the demands of other people. Nevertheless it is certainly possible for future values to outweigh present ones, depending on the situation, since future values do have *some* weight.

Pettigrew allows for several features to be relevant to the weightings assigned to past and future selves. The degree of psychological connectedness between the present self and past and future selves is one such factor.[10] Relatedly, the degree to which past and future selves share one's present values may be relevant. Thus the desires I have in

[9] Trebilcot notices a similar parallel with morality but derives a different conclusion: "the aprudentialist may feel it an unacceptable violation of the value and integrity of the present to spoil it deliberately, regardless of whether doing so will produce greater benefits later on" (1974, 210).
[10] See McMahan (2002) for a similar view; both follow Parfit (1984) in seeing psychological connectedness as especially important in prudential reasoning.

the near future will likely be more weighty than desires in the far future, since my distant future self will be less strongly connected, and share fewer of my present values, than my near future self. Another relevant consideration is that one's present self may have obligations to one's past selves based on sacrifices they have made, which may require extra weight to be given to past selves. For now let us consider a version of Pettigrew's view that includes only psychological connectedness, including shared values, as relevant to the weightings; this will not affect the argument of the next section, which will apply no matter what determines the weightings.

In order to state Pettigrew's view more perspicuously, and to compare it with other views, it will be useful to introduce a notion of *time-relative prudential value*. To get the time-relative prudential value of a moment, we start with the intrinsic value of the moment for the subject, as determined by the simple view. We then multiply that by a fraction corresponding to the degree of psychological connectedness between the subject at the moment being evaluated and the subject at the time to which the prudential value is being relativized. The relevant time, for Pettigrew, is the time of decision. That is, we say that the t-relative value for S of a moment t_n = (the value for S of t_n) times (the degree d of psychological connectedness ($0 \leq d \leq 1$) between S at t and S at t_n). We can use this notion of t-relative prudential value to state Pettigrew's view as follows: an act performed at t by S is prudent iff it maximizes the sum, for all times t_n, of the t-relative value for S of t_n.[11] The view as I've stated it here need not include a commitment to subjectivism about well-being; it is compatible with any substantive view about well-being. In fact, with the commitment to subjectivism removed, Pettigrew's view is very similar to McMahan's Time-Relative Interest Account of the badness of death (McMahan 2002). Of course, only if we presuppose a subjectivist account of well-being will the view be seen as a response to the problem of changing values; otherwise it will merely be a way to account for varying degrees of prudential concern, which need not accompany a change in values.

There are other time-relative views one might hold. For example, there is a view that endorses brute temporal discounting of the future. On this sort of view, a "near-biased" or "impatient" view, future values are discounted merely because, and to the extent that, they are future (Parfit 1983). I won't endeavour to discuss every possible time-relative view here.

III. The Connection Between Prudence and Beneficence

[11] I diverge from Pettigrew's statement of his view in one important way. Pettigrew puts the view this way: "We ought to maximize the subjective expected utility from the point of view of the aggregate of our various selves… our aggregate utilities for a given state are given by a weighted average of our past, present, and future utilities within that state" (Pettigrew 2019, 76). By formulating the view in this way, Pettigrew opens up his view to intrapersonal analogues of objections to average utilitarianism (see Arrhenius et al. 2022 for an overview). It will, for example, favor a long life of torture over a much shorter life of slightly worse torture. My reformulation avoids those problems, but of course faces problems of its own. Thanks to Adriano Mannino for discussion.

Recall Williams's claim: "the correct perspective on one's life is from now." Williams is talking about the agent's perpective on *their own* life. But what is the correct perpective to take on *someone else's* life?[12] If you are making the decision for me, and you are attempting to make good decisions on my behalf, what do you have most reason to choose?

The central claim of this paper is that the answer to this question must be very tightly connected to the first-personal question. Suppose Pat is in need of medical treatment, and is in a state in which he is passing in and out of consciousness. Suppose Kat is Pat's medical proxy. Kat has a special obligation to make decisions that are good for Pat. Pat happens to be unconscious at the moment and unable to make a decision that must be made urgently: choose treatment A or treatment B. Kat knows that it would be more prudent for Pat to choose treatment A than treatment B, if Pat were able to make a choice. It is no more costly for Kat herself to choose A over B. Nevertheless Kat, in her capacity as medical proxy, chooses treatment B. This would be an outrageous dereliction of duty on Kat's part.

You might think that it could be permissible for Kat to choose a treatment that would be imprudent for Pat to choose for himself. For example, suppose Pat would in fact, imprudently, choose treatment B for himself, and Kat knows this to be the case.

[12] Thoma (2023) raises this question regarding the correct attitudes towards *risk* on another's behalf, but not regarding the axiological question I am investigating here about the correct attitudes towards *benefit*.

Perhaps in that case Kat should choose what Pat *would* choose for himself, rather than what Pat *should* choose were he to be prudent. This might be required by considerations of autonomy. It is, perhaps, disrespectful for Kat to impose her own values on Pat, overriding Pat's views about what is best. These considerations are important, but I take it there is a difference between acting beneficently towards someone and acting in a way that respects their autonomy. I am focusing here on beneficent action, not autonomy-respecting action. I am not attempting to provide a general theory of proxy decision-making that incorporates all factors that can affect how one should choose for another. So let us just stipulate that in the example, there is no fact of the matter about what Pat *would* choose. Whichever choice Kat makes, it will not violate Pat's autonomy. It seems clear, given this stipulation, that Kat should choose the option that would be prudent for Pat to choose.[13]

What is the best explanation of Kat's failure? First, let me introduce the notion of an "ideal benefactor." Here is a generic definition of this term: an *ideal benefactor* of S always does what is in the best interests of S. I propose, for now, the following simple principle linking prudence with beneficence:

[13] Thanks to Sean McKeever for raising this issue. Issues of autonomy might raise difficulties for my proposal here, if autonomy factors into well-being; I discuss this issue in the appendix.

Prudence/Beneficence Link: It is prudent for S to choose A over B in circumstances C if and only if S's ideal benefactor would choose A over B for S when S is in C.[14]

If this principle is true, Kat does not act as an ideal benefactor would by acting differently from the way Pat would, were Pat to act prudently. This (along with the stipulated absence of countervailing reasons, including autonomy considerations) is what makes her conduct unacceptable.

The Prudence/Beneficence Link follows from a more general thought about prudence and beneficence: *prudence is just a special case of beneficence* in which the benefactor and the one benefited are the same individual.[15] A theory of prudence provides each of us with a ranking of alternative courses of action that we ourselves might undertake. The Prudence/Beneficence Link says that the ranking may be used by anyone, not just the agent, when acting on the agent's behalf. Denying this link tells us that the ranking is useless to anyone except the agent themselves. It would also create strange dilemmas of the following sort. Suppose we have true theories of prudence and beneficence. The theory of beneficence has implications for the special case in which the agent and the target of benefit are the same individual. If, in such cases, the theory of

[14] While I think this formulation of the principle is at least roughly true, there are complications. To maintain the flow of the paper I have put this discussion in an appendix.

[15] Jeff McMahan explicitly rejects this connection between prudence and beneficence: "While prudence may permit an individual to discount her own future interests for the weakness of certain relations between herself now and herself in the future, morality requires that we not discount the future interests of others" (McMahan 2006, 629).

beneficence were to give different verdicts from the theory of prudence, an agent could find herself in a strange practical dilemma where she would have two true theories giving conflicting guidance about how to treat herself based on the same underlying facts about the production of well-being or the satisfaction of the agent's values. It is hard to see what could resolve such a dilemma. Accepting the Prudence/Beneficence Link ensures that no such dilemma is possible.

Time-relative views such as Pettigrew's are incompatible with the Prudence/Beneficence Link. The source of the problem is that relativistic views are *agent*-relative in a way that would undermine the link. To show the problem more precisely, consider the following slightly more substantive and less neutral principle about ideal benefactors:

Ideal Benefactor Principle: S's ideal benefactor will choose *the best overall life* for S. [16]

We cannot accept this Ideal Benefactor Principle in conjunction with the Prudence/Beneficence Link and a time-relative view. The Present-Aim View requires the agent to completely discount their future goods in their prudential deliberation; Pettigrew's view merely permits this, as well as permitting partial discounts; but the Ideal Benefactor Principle *requires* a benefactor *not* to discount at all. So the

[16] James Snowden endorses this principle, modified to account for risk: "an altruistic donor should prefer an action over an alternative if that action results in higher expected welfare for each beneficiary" (Snowden 75).

Prudence/Beneficence Link is inconsistent with the combination of the Ideal Benefactor Principle and a time-relative view of prudence, even one that permits rather than requires discounting for e.g. psychological connectedness.

The Ideal Benefactor Principle as stated makes beneficence obviously time-neutral, so it is no surprise that a time-relativist about prudence could not accept both it and the Prudence-Beneficence Link. So we should consider ways to formulate a time-relativized principle about ideal benefactors. Perhaps we could say that a benefactor should or could discount the future values of the person they are trying to benefit, just as that person should or could. But it is unclear what would motivate such a discount. Recall that for Pettigrew, the motivation for an agent to discount their future values is the fact that future versions of the agent are relevantly like other people. But that gives the *benefactor* no reason to discount! Present, past, and future versions of me are all actually different people from you, and you need not have any special connection to my *current* self (though in some cases you might). You have no reason to favor my current values, so you should not discount my future values if you are concerned for my welfare.

Here is another thought. I've talked about an ideal benefactor as someone who makes the best choice for S. But if we take time-slices of people seriously, perhaps the notion of an ideal benefactor should be time-relative: an ideal benefactor is someone who makes the best choice for *S-at-t*. The ideal benefactor of S-at-t1 would make a different choice from the ideal benefactor of S-at-t2 if S has different values at t1 and t2. But this merely relocates the problem. If I intend to benefit someone, I will need to choose which time-slice of them I should choose for, but there is no reason to choose for one time-slice over another.

Perhaps the spirit of the time-relative view would tell a benefactor to give priority to the values of the time-slice that exists at the time the *benefactor* is making the decision. But consider an example like the following. Suppose my friend Kim, a fan of the Buffalo Bills, moves to Cleveland. We are out of touch for some time, but I happen to know that because of her move to Cleveland, within a year she will join the Church of the Subgenius, and will change her values accordingly. She will cease caring about sports and will care only about Subgenius-related things. I now can buy her a gift, which she will not receive for one year due to supply chain disruptions. I can buy her a Buffalo Bills jersey or a t-shirt with a portrait of J.R. "Bob" Dobbs, mythical founder of the Church of the Subgenius. Currently she strongly prefers to have the jersey and has no desire for the Bob shirt, and in fact would be appalled by a future in which she wanted to wear such a shirt. But in a year, her preference will be reversed. On the proposal I'm floating, where it is the time of the decision that is relevant, I would seem to have some reason to buy her the jersey, given her current values; but this cannot be right. To see why not, we can just imagine that I don't know whether she has undergone her value conversion yet, and imagine whether it makes sense for me to call her to find out before I buy the gift. Clearly it does not matter.

In sum, the problem for time-relative views is that the motivation for favoring the present does not carry over from agents to benefactors. So there seems to be no way to reconcile such views with the Prudence/Beneficence Link.

IV. World-Relative Prudence

Let us now turn to the other objection to the Simple Time-Neutral View: it gives agents a prudential reason to change their values to ones that are easier to satisfy. Among those who find such a reason objectionable is Rawls, who writes disparagingly about what he calls "bare persons": "Such persons are ready to consider any new convictions and aims… when doing this promises a life with greater overall satisfaction, or well-being" (Rawls 1982, 181).[17] Bricker's complacency pill example draws out the counterintuitiveness of the idea that it is prudent to make such changes to one's values. Dale Dorsey gives a similar example also involving a pill: suppose I currently want nothing more than to be an astronaut, but I can take a pill that will cause me to want to

[17] See also Rosati 1996, 302. I am not sure whether Rawls intends to be making a claim about welfare or reasons for action; perhaps he thinks that a person such as the one he has in mind would indeed make themselves better off by changing their preferences, but that they shoudn't do it anyway (in some sense of "shouldn't"). He seems to be casting doubt on the possibility that such a person could exist at all, but also seems to be trying to give an argument against hedonism. Rosati on the other hand is clearly making a point about well-being, so she presumably would disagree with Dorsey's thought that in these kinds of cases there is prudential reason to do what makes one's life worse.

be a pilot instead. And suppose that since it is easier to be a successful pilot than a successful astronaut, my pilot-desires would be better satisfied, were I to have those desires, than my astronaut-desires would be if I were to have them. (Stipulate that it is not objectively better to be an astronaut than a pilot.) On the simple view, it would be prudent for me to take the pill, because it would make my life better. Dorsey thinks this is wrong; it would not be prudent to take the pill (Dorsey 2021, 212). As with Pettigrew's view, this is because of a concern about prudential alienation. If I don't value becoming a pilot, a theory of prudence that required me to become a pilot anyway would be intolerably alienating; it doesn't seem to make things much better to say that I should just *start wanting to be a pilot* so as to be better able to get what I want. Dorsey puts the concern this way: "In changing one's future prudential orderings, one *necessarily* does not respond to facts about what *is* good, because this action… simply grants the status of *good*… to particular states of affairs" (Dorsey 2021, 217, his emphasis). Taking the pilot pill, according to Dorsey, would not be promoting what is good for me, but *changing* what is good for me, and I have no prudential reason to change what is good for me.[18]

[18] Bruckner (2009) holds a version of the view Dorsey defends here; he thinks prudence is silent about what to do in these kinds of cases. An anonymous referee points out that Dorsey's view is closely linked to the question of the rationality of adaptive preferences; see Elster 1983 and the vast literature that followed, including Sen 1995 and Nussbaum 2001. Defenders of the irrationality of adaptive preference need not hold the view that there is never prudential reason to change one's values.

Dorsey proposes that when making the decision in the astronaut/pilot case, one should completely ignore the future pilot values one would acquire were one to take the pill, because those values are the result of making that decision. Since taking the pill would cause one's values to change, the value-satisfactions that would result from taking the pill do not count in favor of taking the pill. Which values, then, are relevant? "The goods one has prudential reason to promote now are fixed by what one's prudential ordering will be independent of current action" (Dorsey 2021, 221).[19] According to Dorsey, the prudential value of an outcome for S at w is determined by S's values at the closest world to w at which S does not act to change S's values. Dorsey thus rejects an assumption common to all of the subjectivist views discussed so far (and I believe to all extant views of prudence). According to those other views, the prudential reason for an agent to choose an outcome at a world is determined by the values the agent holds *in that world*.[20] Dorsey denies this; he thinks the agent's reason to

---

[19] Dietz (2023) introduces a similar view, but about welfare itself rather than prudence, which he calls "Choice-relative Desire Satisfactionism (CRDS)": "The satisfaction of a desire benefits S, from the perspective of an evaluator M at time t, only if the existence of this desire does not depend on what M does at t" (Dietz 2023, 992). Dietz points out several problems with such a view. There is an important question here that I leave to the side, which a defender of Dorsey's view (or CRDS) needs to answer: what is it for a prudential ordering to be "independent" of current action? Since my main aim here is not Dorsey criticism I will hope to just go along with whatever Dorsey thinks is the correct answer to the question of what counts as the default world; my objection to Dorsey will go through no matter how he answers this question.

[20] See e.g. Bykvist 2006, 279: "No cross-world intervention is allowed: a life should be judged only by the desires one would develop if one were to lead that life." Pettigrew points out in his discussion of Harman (2009) that this same feature is present in so-called "I'll be glad I did it" reasoning: such reasoning "asks you to compare your future utility, should you choose one option, for the world you create by choosing that option and your future utility, should you choose that same option, in the world you would have created by choosing the alternative" (Pettigrew 2019, 227).

choose an outcome is determined by what the agent values in the closest world where the choice does not change the agent's values (which might of course be the actual world depending on what choice the agent makes). Note that in other cases Dorsey's view gives us no discount at all. If, without my doing anything to make this the case, my values would have changed from astronaut values to pilot values, then I do have reason to take my future pilot values into account, and those values should not be discounted at all.

In order to provide a precise statement of Dorsey's view, it will be necessary to introduce a notion of *world-relative prudential value*. For any worlds w and w_m and time t, let us say that the $<w_m,t>$-relative value for S of $<w,t>$ = the extent to which $<w,t>$ matches the values S has at $<w_m,t>$, adjusted for strength. (Note that this understanding of world-relative value incorporates a commitment to subjectivism, since world-relative value is defined in terms of an agent's values; thus unlike Pettigrew's view, Dorsey's cannot accommodate objective theories of welfare.) Then we can state Dorsey's view as follows: an act performed at world w and time t is prudent for S iff it maximizes the sum, for all times t_n, of the $<w_m,t_n>$-relative value for S of $<w,t_n>$, where w_m is the closest world to w at which S does not act at t so as to change S's values. It is worth noting here that Dorsey's view is in a way both world- and time-relative, since w_m is determined by appeal to what happens at time t.

Dorsey's view faces a prima facie difficulty in accommodating the Prudence/Beneficence Link. Suppose you are deciding whether to give me the pilot pill. On Dorsey's view it is prudent for me to to refuse the pill. But my life will be better if I take it; so if the Ideal Benefactor Principle is true, my ideal benefactor would give it to me. So Dorsey cannot consistently endorse the Prudence-Beneficence Link and the Ideal Benefactor Principle. But the Ideal Benefactor Principle is formulated in a way that is congenial to the Simple Time-Neutral View anyway, so this should perhaps not be a surprise. And in this case, an alternative ideal benefactor principle might be plausible. Dorsey could say that just as I have no reason to bring about a future where I have different values, you as my benefactor have no reason to bring about a future where I have different values. Thus he could adopt the following ideal benefactor principle: S's ideal benefactor will, at t, choose the life for S that is best as determined by all of S's prudential orderings, at all times, as they would be if *the benefactor* were not to act at t so as to change them.

Adopting such a principle would, I think, allow Dorsey to retain the Prudence/Beneficence Link. But other problems would remain. Return to Dorsey's pilot example. Stipulate that it would be imprudent, as Dorsey says, for me to take the pilot pill. But suppose you secretly put the pilot pill in my drink, and I unwittingly drink it. In a day's time, it will take effect, and I will want to be a pilot. You then come to believe that you should not have given me that pill. As it happens, you have an antidote that you could secretly give to me, in which case my values will never change. What should you do? Given the ideal benefactor principle I have proposed for Dorsey, you should not give me the antidote. After all, if you were to do so, this would change my future values from what they would be if you didn't now act. But how could it be contrary to beneficence for you to give me the pill in the first place, but also contrary to beneficence for you to give me the antidote afterwards?

Dorsey's account of prudence is committed to a kind of "prudential inertia." By acting imprudently, I can change which future I would subsequently be most prudent to choose, *without changing what those futures would be like for me*. I would change what is prudent for me merely by changing what counts as the "default" option; by taking the pilot pill, I change the default future to the pilot future; by taking the antidote, I change it back. This is an unwelcome feature of Dorsey's view of prudence. The revised ideal benefactor view incorporates this unwelcome feature too, and results in an analogous "beneficial inertia" whereby what counts as benefiting someone depends on what counts as the default option.[21]

V. CONCLUSION

The Simple Time-Neutral View faces an alienation objection; it recommends acting so as to satisfy one's future values rather than one's present values in many

[21] I adapt the name of the phenomenon from Sartorio 2008. See Hawkins 2015 for more skepticism of the idea that there is a prudential problem with changing one's values.

cases. It also faces a direction-of-fit objection; it recommends changing one's values to ones that will be better satisfied. I've argued that views that depart from the Simple Time-Neutral View in order to avoid these objections face serious problems. Time-relative views such as Pettigrew's are incompatible with thinking of prudence as a special case of beneficence; they are incompatible with the Prudence/Beneficence Link. World-relative views such as Dorsey's may be made compatible with the Prudence/Beneficence Link, but will yield the unintended and implausible result that there is such a thing as prudential and beneficial inertia.

How then can we explain the apparent force of the objections? Consider first the alienation objection. The implication that we should sacrifice present values for future ones seems like a problem only if we privilege the present perspective in our decision-making. Clearly we often do this; but I can see no good prudential justification for doing so. The intrapersonal case and the interpersonal case strike me as importantly different. When I sacrifice present values for future ones, the future values for which I am sacrificing are *mine*. When I sacrifice my own values for someone else's, the other person's values are not mine. Diachronic identity matters. Parfittians such as Pettigrew will fundamentally disagree here. But if thinking of matters of diachronic identity in a Parfittian way leads to the problems I have raised, we should perhaps see this as a problem for that way of thinking about diachronic identity.

The specific examples used to generate the objection are often problematic as well. Consider the example Isaacs uses to motivate giving total preference to present utilities (contrary to both Pettigrew and myself):

> Suppose that you currently value the happiness and wellbeing of your family. Nontheless, you know that tomorrow you will undergo a horrifying (albeit temporary) change, and will come to instead value the misery and wretchedness of all those you currently hold dear… What should you do? …[S]ince what your future self will want is so diametrically opposed to what you want, you should feel compelled to prevent your future self from getting what your future self will want. (Isaacs 1078)

The problem here is that the example involves an obvious value change for the worse. I doubt there would be any temptation to privilege an agent's present values if the value change went in the opposite direction. If not, then the example has no probative force; the judgment that the agent should follow their present values is based not on their presentness but on their obvious superiority as values. To test whether present values should count for more, we should consider cases where the value change clearly does not involve a change for the better or worse. In such cases, for my own part, I would not feel any strong pull to weight present values more heavily. [22]

Now consider the direction-of-fit objection. How bad is it to think that we should change our values to ones that the world will better satisfy? I think it is not very bad at all. Again, we need to imagine that this does not involve changing our values to worse ones; this will eliminate the cases that likely seem most objectionable, where e.g. one decides to become a Nazi or a sadist. It seems to me that it is eminently prudent to be open to changing one's values. In fact it seems to me that privileging one's "inertial" values, or the ones we would have were we to keep them the same, is implausibly conservative. We knowingly engage in many activities in full awareness that doing so might change what we value by exposing us to new ideas or experiences (Heathwood 2019, 679). Engaging in such activities can be prudent even if it comes at the expense of satisfying the values we already had or would otherwise have had.

Does this mean we should accept the Simple Time-Neutral View after all? I think not. The Simple Time-Neutral View faces other problems. Notably, it requires us to satisfy our past desires. I lack the space to adequately discuss this problem here. I think the most promising way to handle these problems *in toto* is to reject the assumption that

[22] It is possible that disagreements about what one should do in these cases are really disagreements about whether there are "external" reasons in Williams's sense ("Internal and External Reasons," in Williams 1981). Thanks to an anonymous referee for this suggestion.

value-satisfactions have something directly to do with well-being or prudence. If we instead adopt an experience-based axiology, such as hedonism, we can avoid the problems I've raised here, and also avoid the requirement to satisfy past desires. But fully defending such a view is a much larger task.

APPENDIX: CLARIFYING THE LINK

The simple version of the Prudence-Beneficence Link is, I think, *roughly* true; but we may wish to sharpen it up. [23]

First, there is the fact that there are in fact no ideal benefactors, or even if there are, there need not always be. Relatedly, there are cases in which the agent is unable to act or make a choice (perhaps because they are incapacitated). Both of these sorts of cases would be problems for the Prudence/Beneficence Link as stated. One option here would be to restrict the circumstances of application: *if* an agent has an ideal benefactor, and is able to make a choice, then the prudent agent would choose in the way the ideal benefactor would choose. Alternatively, we could appeal to a hypothetical ideal benefactor, as follows:

[23] Thanks to Krister Bykvist for raising the objections in this appendix and helpfully suggesting revisions. I've largely followed his suggestions.

Prudence/Beneficence Link version 2: If C are circumstances in which S can choose between A and B, then it is prudent for S to choose A over B in C if and only if S's *hypothetical* ideal benefactor would choose A over B for S when S is in C.

Another problem might arise if there are situations where it is important for the agent to make the choice herself. Suppose for example that autonomous choice factors directly into well-being; an agent is better off if they make a choice themselves than if the choice is made for them by someone else. In an extreme case, a choice made for someone by someone else would not benefit that person at all. Perhaps choosing one's spouse is like this. It would be best if the Link were formulated so as to be neutral between theories of well-being. So we might wish to reformulate the Link in terms of preference rather than choice, as follows:

Prudence/Beneficence Link version 3: If C are circumstances in which S can choose between A and B, then it is prudent for S to choose A over B in C if and only if S's hypothetical ideal benefactor would *prefer* A over B for S when S is in C.

Finally, and most interestingly, one might wonder if the bias towards the future poses problems for the Link.[24] This worry arises especially when we introduce the notion of preference into the link as in Version 3 of the Link, since arguably a bias

[24] Thanks to Jamie Dreier and Krister Bykvist for raising this worry.

towards the future will not affect rational *choice* but will affect rational *preference*.[25] Some people report having a bias towards the future when considering themselves, but not when considering other people. For example, in Parfit's "My Past or Future Operations" case, some people think that it makes sense to prefer to have the longer, more painful operation in one's own past than to have the shorter, less painful operation in one's own future, but that it would not make sense to have similar preferences concerning someone else's past or future suffering (Parfit 1983). If this combination of attitudes is rational, then at least sometimes it could make sense for an ideal benefactor and the person they are benefiting to differ in their rational preferences.

Because I cannot hope here (or perhaps anywhere) to sort out the rationality of brute time biases, I am inclined to retreat to a principle that is slightly weaker but still strikes me as interesting. Let us simply take the ideal benefactor of an agent to be one who shares the agent's time bias: if the agent prefers their own pain to be in the past, the ideal benefactor of that agent will also prefer the agent's pain to be in the past. This does not strike me as begging any relevant questions, and the resulting Prudence/Beneficence Link would still be incompatible with the time-relative views discussed in Section II.

---

[25] See e.g. Parfit 1983 and Edidin 1982 on this point; but see Dougherty 2011 for an argument that future bias can affect rational choice. See Sullivan 2018 for an extended treatment of time bias.

References

Arrhenius, Gustaf, Jesper Ryberg, and Torbjörn Tännsjö. 2022. "The Repugnant Conclusion." In Edward N. Zalta & Uri Nodelman (eds.), *The Stanford Encyclopedia of Philosophy* (Winter 2022 Edition), URL = <https://plato.stanford.edu/archives/win2022/entries/repugnant-conclusion/>.

Bradley, Ben. 2020. "A Defense of Temporal Well-Being." *Res Philosophica* **97**: 1-7.

Bramble, Ben. 2020. *The Passing of Temporal Well-Being*. Routledge.

Brandt, Richard. 1972. "Rationality, Egoism, and Morality." *Journal of Philosophy* **69**: 681-697.

Bricker, Phillip. 1980. "Prudence." *The Journal of Philosophy* **77**: 381-401.

Bruckner, Donald. 2004. "Prudence and Justice." *Economics and Philosophy* **20**: 35-63.

Bruckner, Donald. 2009. "Silent Prudence." *Philosophical Explorations* **13**: 349-364.

Bykvist, Krister. 2006. "Prudence for Changing Selves." *Utilitas* **18**: 264-83.

Bykvist, Krister. 2022. "Wellbeing and Changing Attitudes Across Time." *Ethical Theory and Moral Practice*. (Published online July 26 2022)

Dietz, Alexander. 2023. "Making Desires Satisfied, Making Satisfied Desires." *Philosophical Studies* **180**: 979-999.

Dorsey, Dale. 2015. "The Significance of a Life's Shape." *Ethics* **125**: 303-330.

Dorsey, Dale. 2021. *A Theory of Prudence*. Oxford University Press.

Dougherty, Tom. 2011. "On Whether to Prefer Pain to Pass." *Ethics* **121**: 521-537.

Edidin, Aron. 1982. "Temporal Neutrality and Past Pains." *Southern Journal of Philosophy* **20**: 423-431.

Elster, Jon. 1983. *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.

Harman, Elizabeth. 2009. "'I'll Be Glad I Did It' Reasoning and the Significance of Future Desires." *Philosophical Perspectives* **23**: 177-199.

Hawkins, Jennifer. 2019. "Well-Being, the Self, and Radical Change." *Oxford Studies in Normative Ethics* **9**: 251-70.

Heathwood, Chris. 2019. "Which Desires are Relevant to Well-Being?" *Nous* **53**: 664-688.

Hersch, Gil. 2022. "The Usefulness of Well-Being Temporalism." *Journal of Economic Methodology*. DOI: 10.1080/1350178X.2022.2129736

Isaacs, Yoaav. 2020. "The Problems of Transformative Experience." *Philosophical Studies* **177**: 1065-1084.

Kauppinen, Antti. 2015. "The Narrative Calculus." *Oxford Studies in Normative Ethics* **5**: 196-220.

McMahan, Jeff. 2002. *The Ethics of Killing*. Oxford University Press.

McMahan, Jeff. 2006. "Paradoxes of Abortion and Prenatal Injury." *Ethics* **116**: 625-655.

Nussbaum, Martha. 2001. "Adaptive Preferences and Women's Options." *Economics and Philosophy* **17**: 67-88.

Parfit, Derek. 1983. "Rationality and Time." *Proceedings of the Aristotelian Society* **84**: 47–82.

Parfit, Derek. 1984. *Reasons and Persons*. Oxford University Press.

Paul, L.A. 2014. *Transformative Experience*. Oxford University Press.

Pettigrew, Richard. 2019. *Choosing for Changing Selves*. Oxford University Press.

Rawls, John. 1982. "Social Unity and Primary Goods." In Amartya Sen and Bernard Williams (eds.), *Utilitarianism and Beyond* (Cambridge University Press), pp. 159-185.

Rosati, Connie S. 1996. "Internalism and the Good for a Person." *Ethics* **106**: 297-326.

Sartorio, Carolina. 2008. "Moral Inertia." *Philosophical Studies* **140**: 117-133.

Sen, Amartya. 1995. "Gender Inequality and Theories of Justice." In Martha Nussbaum and Jonathan Glover (eds.), *Women, Culture, and Development: A Study of Human Capabilities* (Oxford: Clarendon Press), pp. 259-273.

Snowden, James. 2019. "Should We Give to More Than One Charity?" In Hilary Greaves and Theron Pummer (eds.), *Effective Altruism: Philosophical Issues* (Oxford University Press), pp. 69-79.

Strotz, R.H. 1973. "Myopia and Inconsistency in Dynamic Utility Maximization." In M.J. Farrell (ed.), *Readings in Welfare Economics* (Palgrave), pp. 128-143.

Sullivan, Meghan. 2018. *Against Time Biases*. Oxford University Press.

Thoma, Johanna. 2023. "Taking Risks on Behalf of Another." *Philosophy Compass*, e12898. https://doi.org/10.1111/phc3.12898

Trebilcot, Joyce. 1974. "Aprudentialism." *American Philosophical Quarterly* **11**: 203-210.

Ullmann-Margalit, Edna. 2006. "Big Decisions: Opting, Converting, Drifting." *Royal Institute of Philosophy Supplement* **58**: 157-72.

Weirich, Paul. 1981. "A Bias of Rationality." *Australasian Journal of Philosophy* **59**: 31-37.

Williams, Bernard. 1981. *Moral Luck: Philosophical Papers 1973-1980* (Cambridge University Press), pp. 1-19.