Journal of the American Philosophical Association (2017) 412–429 © American Philosophical Association
doi: 10.1017/apa.2018.1

Must Adaptive Preferences Be Prudentially Bad for Us?

ABSTRACT: *In this paper, I argue for the counterintuitive conclusion that the same adaptive preference can be both prudentially good and prudentially bad for its holder: that is, it can be prudentially objectionable from one temporal perspective, but prudentially unobjectionable from another. Given the possibility of transformative experiences, there is an important sense in which even worrisome adaptive preferences can be prudentially good for us. That is, if transformative experiences lead us to develop adaptive preferences, then their objects can become prudentially better for our actual selves than the objects of their nonadaptive alternatives would now be. I also argue, however, that the same worrisome adaptive preferences might still be* prospectively *prudentially objectionable: that is, our pretransformation selves might be prudentially better off undergoing a nonadaptive alternative transformative experience instead. I argue that both claims hold across the range of the most broadly defended accounts of well-being in the literature.*

KEYWORDS: adaptive preference, prospective good, retrospective good, transformative experience, well-being

In this paper, I consider two seemingly conflicting claims about adaptive preferences: (a) they are prudentially harmful to persons who hold them, and (b) they can accurately capture persons' prudential good. Despite the seemingly fundamental conflict, I argue that these claims might both be true of the same adaptive preferences—each claim might simply be true of those preferences from a different temporal perspective. That is, for the same adaptive preference, it might be the case that before the adaptive preference is formed, the agent would be made prudentially worse off by forming the adaptive preference than by forming a nonadaptive alternative. However, after the adaptive preference has been formed, the agent would be made prudentially better off by satisfying the adaptive preference than by acting in accordance with whatever the nonadaptive alternative would have recommended. I argue that the two claims hold both broadly and robustly. They hold broadly insofar as they can be made sense of by all theories

For generous and helpful feedback on various versions of this paper, I am grateful to Stephen Campbell, Jonathan Herington, Scott Tanona, Valerie Tiberius, audiences at the Australian National University, participants at the second annual Kansas Workshop on Well-being, and two anonymous referees for this journal. I am also grateful to the Murphy Institute at Tulane University for the Faculty Fellowship that allowed me to complete this paper surrounded by such wonderful philosophers and such an exceptional city.

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

of well-being most frequently defended in the literature. And they hold robustly insofar as the contributions to well-being made by firmly entrenched adaptive preferences can remain even after adaptive pressures are removed.

My argument proceeds as follows. In section 1, I define adaptive preferences and clarify my project. In section 2, I introduce the concept of transformative experience and explain how the relationship between it and adaptive preference allows the two temporal claims to both be true of the same adaptive preference. In section 3, I show how adaptive preferences can be prudentially good for persons posttransformation across a range of broadly defended theories of well-being. In section 4, I show how adaptive preferences can simultaneously be bad for pretransformation persons across the same set of theories. And in section 5 I conclude by noting that we must take these two different judgments about prudential value equally seriously.

1. What are Adaptive Preferences and What are We Asking?

In this paper, I adopt the following definition of adaptive preference:

> **Adaptive preference:** A favoring of an option, *O*, developed in response to a constrained option set that includes *O* as a member.

This initial definition is intentionally thin because, as we will see, scholars writing on adaptive preferences tend to understand the concept quite differently, and I do not want to beg the question between them. But despite the definition's ecumenicism, note that it is both narrower and broader than it could be.

It is narrower because a preference only counts as adaptive if it was developed *in response* to an option set's limitations. That is, the limitations must play a constraining role in the adaptive preference's development. If I prefer some relatively unambitious option within my limited set, but would have preferred the same option even if many others had been available, then my preference is not adaptive on our definition. Additionally, the preferred option must be a *member* of the limited set of options. This is noteworthy because the stark narrowness of an option set can sometimes be exactly the feature that causes us to desire more. Consider a female student in the 1950s who wants to attend Harvard precisely because she resents the university's unjust exclusion of women. Although this preference is formed in response to a limited option set, it does not count as adaptive because the preferred option lies outside the set of accessible options.

The definition is also broader than it could be because it includes more than preferences proper. Strictly speaking, a preference is a three-place predicate: *A* prefers *B* to *C*. An individual preference, then, tells us very little about a person unless it is part of a (relatively) full ordering. In the literature on adaptive preferences, however, it is commonplace to discuss things akin to preference that tell us, on their own, more about what matters to a person. These include desires, values, commitments, and some beliefs and features of a person's character. Since complete preference orderings are rare and since we presumably care about well-being for the

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

sake of whole persons, a focus on preferences proper will be underinclusive. When I use the term, I accordingly make implicit reference to this larger set of nearby concepts.

Initial definition in hand, we can turn to our primary question: Must adaptive preferences be prudentially bad for us? I use 'prudential good' interchangeably with 'well-being' and mean by it what is classically meant: what makes a life good for the person living it, rather than for others. Prudential goodness, then, is distinct from other kinds of value (moral, aesthetic, etc.) that a life might have, and a life might be quite prudentially good for a person even if it is wanting from these other perspectives. But while the primary question asks whether adaptive preferences compromise our well-being, note that on most accounts of well-being, *simply having* an adaptive preference will have no effect at all. Rather, *satisfying* an adaptive preference could compromise well-being, or the *object* of a person's adaptive preference could be prudentially bad for her. However, since these formulations are cumbersome, I will often use the language of adaptive preferences compromising well-being as shorthand for them.

We might think that the answer to our primary question is an obvious 'yes'. After all, the lion's share of work on adaptive preferences concerns the ways in which socially-limited desires can fail to capture a subject's actual interests. Virtually all classic references to adaptive preferences are of this sort: they concern slaves content with their captivity, women who endorse dangerous and inaccessible beauty norms, men who endorse norms of masculinity that preclude emotional expression, etc. Indeed, many accounts of adaptive preferences build this prudential badness directly into their definition. Consider Anita Superson's account, on which adaptive preferences necessarily benefit third parties to the detriment of the one who holds those preferences—and on which adaptive preferences almost always conflict with a person's own more authentic desires to advance her own welfare (Superson 2005). Or consider Serene Khader's account or Martha Nussbaum's early account, each of which count as adaptive only those preferences that frustrate basic human flourishing (Khader 2011; Nussbaum 1992, 2001).

Others, however, will answer 'no'. Donald Bruckner, for instance, argues that given the cost of dissatisfaction to human well-being, it is often more rational to pursue adaptive preferences in order to enhance well-being than to pursue nonadaptive alternatives (Bruckner 2009). And Dale Dorsey argues that initially adaptive preferences can over time become 'deep'—that is, they can become integrated into a person's identity in a way that gives them normative authority over her well-being (Dorsey 2013).

We might think that the point of this paper is simply to determine which group is right—can adaptive preferences sometimes be prudentially good for us, or can they not? What I will argue, however, is that for at least some adaptive preferences, both groups might in fact be right. The seeming necessity of choosing between them arises from mistakenly thinking that there is only one relevant temporal perspective from which the question might be asked. What I will go on to argue is that it will sometimes be the case that a negative prudential judgment about an adaptive preference is prospectively appropriate, while a positive prudential judgment is appropriate retrospectively. That is, before the adaptive preference is formed, the

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

agent would be made prudentially worse off by forming the adaptive preference than by forming a nonadaptive alternative, but after the adaptive preference has been formed, the agent would be made prudentially better off by satisfying the adaptive preference than by acting in accordance with whatever the nonadaptive alternative would have recommended.

Before I make that argument, however, let me distinguish the project from two nearby but less ambitious ones. First, I am not merely aiming to show that adaptive preferences can be prudentially beneficial in the limited sense that they reduce frustration. I take it that virtually anyone writing on adaptive preferences will grant this. If Aesop's fox cannot reach the grapes, his options are to eat no grapes and feel frustration or to eat no grapes and feel no frustration because he has convinced himself that they are sour. Given that he gets no grapes either way and that frustration decreases his well-being, the fox is clearly prudentially better off with the second option. Even Superson, who builds prudential badness directly into her account, recognizes that women in sexist societies will derive *some* limited prudential benefits from happily complying with sexist norms (Superson 2005: 110–11).

But note that the question for the fox is really *which action* he ought to take given the options accessible to him, not a question of *which options* it would be prudentially best for him to have. If everyone agrees that avoiding frustration has prudential value, then our inquiry should focus not on what persons should do given limited options, but on what would be prudentially best for persons if we could make other external options available. That is, we should be interested in what is prudentially best given facts about persons, not facts about external obstacles.

Note that this distinction is different from the distinction between basic and instrumental goodness. That distinction holds that something is basically good if it is good in and of itself, not because of its relationship to other goods. Something is instrumentally good if it provides access to something that is basically good. Thus, sexual pleasure might be basically good, and sexual activity might be good insofar as it provides pleasure. I am instead interested in the question of whether external changes would be prudentially worth making, given who a person is (or, as we will see, is capable of becoming). Imagine the difference between two people, neither of whom has a preference for engaging in sexual activity: a person who has only ever had inept lovers, and a person who is asexual. It would be prudentially good for the former, but not the latter, to change the world by providing her with sexually skilled, attentive partners. This is quite different from asking whether sexual pleasure is basically or instrumentally prudentially good. Holding only facts about the fox—or who the fox could become—fixed, the grapes are still prudentially valuable because he really would find them delicious if he ate them. The interesting version of our question, then, is this: can the *object* of an adaptive preference be one that it is prudentially good or best for a person to achieve in light of facts about who she is or could become?

Second, I am not merely aiming to show that the category of adaptive preferences is so wide that some might prudentially benefit some people. Imagine that as a teenager, I would have formed a preference for regularly using heroin if I had known

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

how to get some. Because I did not know that, however, I decided to apply myself in school. On our definition, this is clearly an adaptive preference: studying hard was within my limited option set, and I would not have preferred it if my set had also included heroin use. But presumably anyone will grant that my adaptive preference is in this case prudentially far better for me than the alternative.

What I aim to show is that the objects of a narrower, intuitively worrisome set of adaptive preferences can be prudentially good for persons, at least from one temporal perspective. Different philosophers will again define this narrow set differently, and I offer the following as an incomplete list of plausible limits. Adaptive preferences might be worrisome only if: (1) an agent lacked objectively better alternatives when forming them, or (2) she lacked alternatives that she would have preferred if she had been more informed, or (3) the object of the adaptive preference itself lacked objective value or would not have been preferred if the agent had been more informed; furthermore, superior options might need to have been available (4) at some particular point in the preference's formation or (5) all along, and those options might have to have been preferred (6) in the real world or (7) only in some idealized counterfactual world. Indeed, when we turn to the feminist literature on adaptive preferences, we should also recognize the significantly more worrisome set that arises from (8) significant deprivation, (9) domination, or (10) injustice. I take no stand on the precise definition of this intuitively worrisome set because nothing in my argument hangs on a precise definition. It should capture, however, the classic cases referenced above: content slaves, dangerous but endorsed norms of femininity and masculinity, etc. It is these intuitively worrisome adaptive preferences—of both the more and less worrisome sort—about which my claim is meant to hold.

I should note that the claim I defend in this paper—namely, that the same intuitively worrisome adaptive preferences can be both prospectively prudentially harmful and retrospectively prudentially beneficial—should ultimately be compatible with at least some current work on adaptive preferences. The work done here adds to the literature not by proposing an alternative to current accounts of adaptive preferences, but rather by identifying a currently unrecognized position that other theorists ought to endorse alongside their own. Those who focus on the harmfulness of adaptive preferences ought to recognize that some firmly entrenched adaptive preferences make real prudential claims even if we have strong prudential reason to keep those preferences from being developed in the first place. And those who defend the prudential value of adaptive preferences ought to recognize that there can be real prudential reason for discouraging the development of some adaptive preferences—even when those preferences might genuinely benefit us once they have been developed.

The position is worth developing at length for two reasons. First, we need an explanation of how the position can be viable—that is, no one has as yet given an account of how two such seemingly conflicting claims about the prudential goodness of adaptive preferences can be rendered compatible. And second, the position is worth developing because, when fully articulated, it offers surprisingly broad and robust conclusions about the extent to which adaptive preferences can benefit persons prudentially. The conclusions are surprisingly broad because—as

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

I will show—they hold for all of the most commonly defended theories of well-being. And the conclusions are surprisingly robust because—as discussed above—they do not show that adaptive preferences benefit us merely by allowing us to resign ourselves to a world we cannot change. Rather, those preferences can become such a central part of who we are that it might no longer be prudentially better for us to change the world even if we could.

2. Transformative Experience and Adaptive Preference

To explain how adaptive preferences can be both prospectively prudentially harmful and retrospectively prudentially beneficial, I appeal to the phenomenon of transformative experience. By 'transformative experience', I mean an experience from which one emerges with a substantially altered set of beliefs and values. Such experiences leave a person with a significantly different character, preference set, and sense of self (Paul 2014; note that my use of the term is broader than Paul's and emphasizes personal rather than epistemic transformation). While transformative experiences are only one sort of event that can lead to this kind of personal transformation and while it is strictly personal transformation that provides our explanation, I focus on transformative experience to emphasize the frequency with which personal transformation depends on the particular and often accidental experiences and circumstances the world presents us with. And as we will see below, my understanding of an 'experience' is broad enough to include many different kinds of personal transformation. Recognizing this, what matters for our purposes is that—without taking a stand on metaphysical questions of personal identity—we should understand persons to emerge, at least sometimes, from such experiences as different people in a meaningful and significant sense.

I say 'at least sometimes' because the changes involved have at least two dimensions, each a matter of degree. First, changes can be more or less mutable. That is, persons will be more or less able to reverse a transformation or undergo new transformations in the same area. For instance, if someone becomes violently ill the first time she drinks alcohol and consequently develops a strong distaste for gin, that distaste might decrease over time or might persist throughout life. Second, the changes involved might be more or less fundamental. A distaste for gin and a love of one's child might be equally immutable, but the latter will almost always be more central to a person's identity. While we all undergo various transformations, some will anchor and organize our fundamental sense of self in ways that others will not. Note that it follows from these two features of transformative experiences that not all transformations will be equally available to all persons. When a feature of a person is less mutable and more fundamental, it will often be difficult—perhaps even impossible—for her to be transformed in that area.

But which experiences are transformative? While the most obvious cases are momentous, unchosen, and discrete—think road-to-Damascus conversions (Ullmann-Margalit 2006)—none of these features are necessary. Indeed, one of the most important transformative experiences that many people will have is that of becoming a parent, and this transition is often a matter of considered choice

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

(Paul 2014). And many chosen transformative experiences occur much more slowly than this one. As creatures with a finite life span, we must prioritize our goals, and attractive goals are often incompatible. When goals conflict, we must make hard choices, and doing so will often require settling on some goals and giving others up, at least temporarily (Goodin 2012). But significant goals are rarely achieved immediately, and as we progress, we go through much slower and more incremental transformative experiences. Consider the young idealist who enters politics and slowly becomes cynical. While this experience is incremental and chosen, it involves fundamental changes to our politician's character and values. She has changed since she settled on her goals, and she would have become different still had she settled on others.

Indeed, we might even fail to recognize we are undergoing a transformative experience. This is plausibly our politician, who fails to appreciate the extent of her cynicism until it is firmly rooted. Edna Ullmann-Margalit describes such cases as 'drifting', where one makes one's big decisions 'conscious of their being decisions, but not conscious of their being big' (2006: 169). In such cases, it is often 'only in retrospect that it can be seen how a particular series of such incremental steps . . . had been all-important in transforming the future shape of [a person's] life and of their personality' (2006: 169–70). Indeed, the most orienting preferences in our lives often result from drifting. Think of your romantic preferences, your preferences for living in the city or the country, or the height of your ambitions. While we can often retrospectively point to experiences that caused these preferences, it is often unclear when those experiences occur how formative they will be.

What, then, is the relationship between adaptive preferences and transformative experiences? Insofar as living with a limited option set can be a transformative experience, adaptive preferences can be the result of transformative experiences. Imagine I want to be a lawyer, but cannot afford the required education. Instead, I reluctantly apprentice with a stone mason. If I had become a lawyer, I would have valued it, but as my apprenticeship continues, I begin to believe that lawyers profit from others' misfortune and that working with my hands is honorable and desirable. Now, I may have consciously convinced myself that masonry is desirable, or I may have simply drifted into caring about it. The change may have occurred instantaneously or gradually. It may be more or less mutable and more or less fundamental. But two things are true: that I would not have become a stonemason if I could have been a lawyer and that I begin my apprenticeship looking down on masonry and end it valuing it highly. While the former fact establishes my preference as adaptive, the latter shows that I have been genuinely transformed. This relationship generalizes: Even if I initially endorsed *P* only because *Q* was not available, living a life of endorsing and pursuing *P* might cause *P* to centrally inform who I am and what I care about.

One might object that the preference change involved in adaptive preference is very different from that involved in transformative experience. According to the objection, in cases of transformative experience one is exposed to new values, and one's preferences, character, viewpoint, etc. shift in recognition. In cases of adaptive preference, however, one settles for what one can get in order to avoid frustration, rather than responding to any genuine value in the newly preferred object. If this is

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

right, then there seems to be little if any overlap between adaptive preferences and transformative experiences.

I reject, however, this characterization of adaptive preferences. Consider two examples of intuitively worrisome adaptive preference: in one, a woman prefers eating only after her husband has finished although she is often left with too little; in the other, a young woman is glad to have undergone female genital cutting although the procedure was painful, and she now lacks access to much sexual satisfaction. In each case, the women's preferences are plausibly grounded in values they can access: the woman who eats last might pride herself on her self-control and self-sacrifice, while the second woman might value ritual, community belonging, and sexual purity. If we grant that these women might respond to value, then we should also grant that their preferences are candidates for transformative experience. To be sure, it might have been unjust and regrettable that they were exposed to these particular values, and if they had encountered values of gender equality or sexual expression instead, then they might have been differently transformed—but the whole point of transformative experience is that one is transformed in response to one's *actual* experiences. And these women, having encountered the values that they did, could well be transformed by them in deep and meaningful ways even if justice might have required different transformations.

Note that the points about mutability and fundamentality also apply to adaptive preferences. As many committed feminist women attest, it can be virtually impossible to stop applying to one's own body the rigid beauty norms that one has explicitly rejected. And as many men attest, expressing emotion freely can remain a challenge even after years of trying to do so. Similarly, if we grant that adaptive preferences can develop in response to genuine values, then we should also recognize that those values can become fundamental to one's sense of self.[1] Self-control and self-sacrifice, for instance, might become the kinds of things that animate every area of a person's life. What this means, then, is that just as not all transformations are equally available to all persons, not all adaptive preferences are equally voidable. Even if all adaptive preferences are initially *avoidable*, some adaptive preferences will be the kinds of things that we are stuck with once they have developed—either because they are sufficiently immutable that we are not psychologically capable of rejecting or altering them, or because they have become so fundamental to our sense of self that we are unwilling to do so even given the robust opportunity.

The relationship between adaptive preference and transformative experience, then, is what explains how some adaptive preferences can be both prudentially good for us and bad for us: while they may be good for one temporal version of ourselves (that is, the posttransformation version), they may also be bad for

1 One might object that if a preference is sufficiently endorsed so that it will be maintained even without a limited option set, it is no longer adaptive. For our purposes this kind of disagreement is merely semantic. Whether one calls such preferences 'adaptive' or 'formerly adaptive' does not affect my claim that, once firmly established, such preferences can make robust contributions to a person's well-being. Similarly, even if such a preference's maintenance is no longer adaptive, the preference's existence is still causally explained by its adaptive origins—and as we will see, adaptive origins can remain sufficient to render a preference prospectively prudentially harmful.

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

another temporal version of ourselves (that is, the pretransformation self). I argue in the next section of the paper that the first half of this claim holds across the range of the most commonly defended theories of well-being in the literature. In the final section, I argue that the second half of the claim holds across the same range of theories. To reiterate, the claim is not that adaptive preferences will alter what is basically good on each theory of well-being—this would be too high a bar since only on preferentist accounts can preferences ever wholly determine basic prudential goodness. Rather, the claim is that adaptive preferences can influence prudential goodness in a robust, person-affecting way: that is, adaptive preferences can lead to the kind of deep, fundamental changes in persons that determine the particular objects of their good across a range of ways in which we could change the world around them. If well-being theorists or political philosophers genuinely care about achieving good lives for persons, then this kind of person-affecting change should be at least as interesting as changes in basic goodness.

3. Adaptive Preferences and Prudential Benefits to Posttransformation Selves

3.1 Hedonist and Desire Satisfaction Accounts

We begin with the easiest case: hedonist and desire-satisfaction accounts of well-being. Hedonist accounts hold that only the sensation of pleasure or happiness contributes to well-being, while simple desire-satisfaction accounts hold that one's well-being consists in the satisfaction of one's actual desires. Since hedonist accounts grant that pleasure may have many sources and desire satisfaction accounts grant that desires may have many objects, transformative experiences can straightforwardly allow the objects of adaptive preferences to contribute to persons' well-being. It must merely be the case that those adaptive preferences' objects be possible objects of genuine desire or pleasure. And this can often clearly be so. Imagine that I originally desired to become a philosophy professor, but upon realizing that I could not then adjusted my expectations and desired to become a high school teacher instead. I could clearly come to take genuine pleasure in my days of teaching and mentoring younger students or genuinely come to desire to succeed in those areas. And if I do, then the object of this particular adaptive preference could clearly contribute to my well-being on hedonist or desire-satisfaction accounts, respectively: being a teacher is not just a consolation prize that helps me to avoid frustration—rather, it is now what I genuinely desire and take pleasure in.

But what of the more worrisome set of adaptive preferences, that is, those involving deprivation, injustice, and domination? Consider a classic case raised in the literature: the dominated housewife who desires to subsume her own needs and projects to her domineering husband's. While readers might be skeptical that our housewife actually desires or takes pleasure in catering to her husband's whims, this skepticism may not be warranted. While satisfying this desire might contribute to the frustration of other actual desires and while the pleasure achieved might

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

be outweighed by pain and fatigue caused by the same circumstances, the original pleasure or desire should still count as at least an instance of pleasure or desire. And indeed, transformative experience might even limit that frustration, pain, and fatigue. Remember, we emerge from transformative experiences with significantly different characters, values, and preferences. If the experience of becoming a dominated housewife is sufficiently transformative, then many of our housewife's other desires and sources of pleasure and pain might fall into line to prevent significant conflicts. In this case, the object of her adaptive preference will be a genuine source of well-being, and she will have changed sufficiently so that few other sources of ill-being will detract from it.

One who remains skeptical of this answer, however, could instead adopt modified versions of these accounts that seem to prevent preferences like the dominated housewife's from contributing to her well-being. I consider L. W. Sumner's modified hedonist account and Peter Railton's informed variation on desire satisfaction accounts. Ultimately, however, I argue that neither gives us sufficient reason to think that satisfying adaptive preferences cannot contribute to well-being.

3.2 An Authentic Life Satisfaction Account

According to Sumner's modified version of hedonism, well-being should depend on the broader category of life satisfaction, rather than on the more limited category of pleasure. But Sumner recognizes that a focus on life satisfaction taken alone might lead to counterintuitive judgments about well-being: namely, that persons like our dominated housewife are prudentially benefitted by the circumstances that they have been socially conditioned into accepting (Sumner 1996: 163–71). Since Sumner is committed to tying a subject's well-being to her own judgments, he aims to correct for this problem by adding an authenticity requirement to his account. This requirement rules out the prudential contribution of nonautonomous life satisfaction, but not because the agents' 'values are objectively mistaken.' Rather, such satisfaction is ruled out because non-autonomous agents 'have never had the opportunity to form *their own* values at all' (Sumner 1996: 166, italics added). That is, the authenticity requirement calls satisfaction like the dominated housewife's into question by suggesting that seemingly sincere endorsement in intuitively worrisome cases is often based in social manipulation rather than in the person's own agency. After all, the 'insidious aspect of social conditioning is that the more thorough it is the less its victims are able to discern its influence on their judgments about their lives' (163).

The authenticity requirement, then, seems to be aimed at foreclosing the possibility that intuitively worrisome adaptive preferences can be prudentially good for us. But even if it succeeds for some adaptive preference tokens, it does not show why this must be true for the entire adaptive preference type. While Sumner does not offer an account of autonomy with which to judge particular instances of life satisfaction, his illustrative examples suggest that at least some types of nonautonomously formed preferences are consistent with later autonomous endorsement. For instance, he discusses a young man raised in a religious community designed to inculcate some values by preventing exposure to

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

others. Although his values are not autonomously formed, Sumner holds that if the young man goes to university, critically evaluates them, and then accepts them, those values now count as authentically endorsed.

While Sumner does not explicitly state why this is so, the explanation seems to relate to the intimate connection between the young man's values and identity. The opportunity to evaluate the values critically, both reinforces and makes obvious that tie. And one way of understanding that tie is to say that being raised in his religious community was a transformative experience that—despite involving the development of nonautonomous and plausibly adaptive preferences—fundamentally altered his character and values. Even if he did not initially have the chance to form his own values, he later had the opportunity to reaffirm them and endorse them as genuinely his own. Thus, on Sumner's account, transformative experiences still seem to make it possible that adaptive preferences can contribute to our well-being. And for any life satisfaction account that is explicitly content-neutral, this should be true at least in principle even of preferences involving domination or injustice.

3.3 An Ideal-Desire Account

Peter Railton's ideal-desire alternative to simple desire satisfaction is motivated by the idea that our 'subjective interests'—that is, the objects of our actual desires— 'frequently reflect ignorance, confusion, or lack of consideration, as hindsight attests' (Railton 2003: 10). Accordingly, he thinks our well-being should be associated not with getting what we actually want, but with what we *would* want if we corrected for these sources of error by idealizing our desires. More precisely,

> an individual's good consists in what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality. (2003: 54)

Some adaptive preference tokens that added prudential value on a simple desire satisfaction account—especially those involving deprivation or domination—will surely fail to receive endorsement from an agent's idealized self. For instance, while the dominated housewife's actual desires might be harmonized sufficiently for most to be satisfied, the idealized version of herself might still want her to leave her husband and develop new desires for independence and equality.

Again, however, this need not be true of the entire adaptive preference type—or even of all those tokens that favor injustice, deprivation, domination, etc. Indeed, an ideal-desire account might offer good reason to think that some adaptive preferences will capture our good *better* than their nonadaptive alternatives. While Railton does not offer a formula for determining what our idealized selves would want, he does note that their desires will be grounded in the 'reduction basis' of our

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

interests: namely, in the facts about an individual's circumstances and constitution that determine, among other things, the person's existing tastes and ability to acquire new ones, the consequences of various courses of action, etc. (2003: 11). If the actual person cannot appreciate or achieve the things that her idealized self wants for her, it is very hard to see how those things could make her life better for her. And since the idealized self seeks to identify those things that make the actual self's life a good one *for her* to live, the idealized self must take into account facts about what the actual self is like and what she may become.

Consider the grounds on which my idealized self might determine whether it would be better for me to be a professional dancer or a doctor. Imagine that I am made anxious by insecurity and do not like change. Knowing these things about my actual self, my idealized self will likely prefer that I desire the life of a doctor (a profession that is always in demand and can be carried on into old age) rather than the life of a dancer (where work is very scarce and can be physically undertaken only by the young). Given my particular reduction basis, being a doctor seems to be better for me on an ideal-desire account than being a dancer.

We can see that the same could be true even of adaptive preferences that favor domination or deprivation. The idealized housewife may *not* prefer that her actual counterpart leave her husband for a new and egalitarian relationship, precisely on the grounds of her reduction basis. Once again, transformative experience here plays a significant role. All persons undergo some transformative experiences, but as we saw above, not all transformative experiences are open to everyone. And part of any given individual's reduction basis will be the set of transformative experiences that is personally available *to her*. Our dominated housewife, for instance, may or may not be psychologically able to undergo a transformative experience that ends with her valuing her independence. And if she is not, then her idealized self will not want her actual self to desire or pursue independence. If the transformation is not psychologically possible, then her pursuit and attainment of independence would be likely to induce anxiety, depression, and loneliness. In this case, it is easy to see why her idealized self would not want her to want independence for the sake of her own well-being. Thus, informed-desire accounts like Railton's show us in an especially clear way why transformative experiences can make the objects of some adaptive preferences prudentially better for our actual selves than their nonadaptive counterparts.

3.4 Perfectionist and Objective List Theories

Finally, we turn to perfectionist objective list theories, where the claim that the objects of adaptive preferences can be prudentially good for us will be least initially plausible. Strictly speaking, perfectionist theories and objective list theories are distinct. While perfectionist theories offer an explanation that unifies the items on their list, objective list theories simply offer a set of objective goods without an explanation for their inclusion. This difference, however, is not relevant for my present purposes, and accordingly I treat theories of these types together. For our purposes, perfectionist objective list theories hold that there is some set of goods that perfect human nature and that are therefore necessarily prudentially good for

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

human beings. In Thomas Hurka's words, certain properties 'constitute human nature or are definitive of humanity—they make humans humans. The good life [then] develops these properties to a high degree or realizes what is central to human nature' (1993: 3). Importantly, this good life does not depend on individuals' reactions to it. Rather, 'it is an ideal people ought to pursue regardless of whether they now want it or would want it in hypothetical circumstances, and apart from any pleasures it may bring' (17).

Since the goodness of an object here depends on the nature of the object rather than on the person for whom it is good, it may be hard to see how preferences—including adaptive preferences—can have any effect on what is good for a person. However, on any plausible objective list theory, they will. Consider first a strict objective list theory, where objects are good for a person whether or not she wants them. I take it that such a theory will only be a plausible account of full prudential goodness if it includes (and indeed, weights relatively heavily) an entry like happiness, pleasure, satisfaction, etc. And while these things may be basically good for persons whether or not they are desired, whether particular persons *experience* them at all will depend on what those persons desire. Next, consider a hybrid objective list theory like Shelley Kagan's, according to which an object can contribute to a person's prudential good only if it both has the correct objective nature *and* is desired, valued, etc. by the agent (Kagan 2009). Those who are drawn to hybrid objective list theories tend to be so drawn because they think that objects' natures matter for prudential goodness, but they also find it counterintuitive that 'something can be (directly and immediately) good for me though I do not regard it favorably, and my life can be going quite well despite my failing to have any positive attitude towards it' (Sumner 1996: 38). Accordingly, hybrid objective list theorists make desire for objectively valuable objects a direct condition of their prudential goodness.

Note that for either version of objective list theory, desire plays a significant role in the extent to which persons experience prudential goodness. And when this is true, transformative experiences can change the objects that contribute to well-being. For the hybrid objective list theorist, those experiences do so whenever a person is transformed from enjoying to not enjoying any object on the list (or vice versa). And for the strict objective list theorist, they do so whenever a person's desires are transformed in ways that cause her to take happiness, pleasure, or satisfaction in different objects. If we have a variety of entries on our list or hold that persons may take pleasure in a variety of things, then such transformations are very plausible. For instance, consider a drafted solider who is transformed through her service into someone who values camaraderie and heroism. If she took little satisfaction in her previous life, then this transformation makes her better off given her new circumstances. On both a strict and a hybrid objective list account it does so because she now has opportunities for satisfaction that she previously lacked—and on a hybrid objective list account she is furthermore better off because she can now access any basic goodness inherent in camaraderie and heroism themselves. And note further that adaptive preferences could involve such changes. Since our solider above was drafted, let us assume that she did not initially approve of military service for its own sake. Despite the role played by her limited option set, however,

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

either objective list theorist should hold that the soldier's successful preference change counts toward her well-being. Both should do so insofar as she experiences real satisfaction, happiness, or pleasure in bravery and camaraderie, and hybrid objective list theorists should further do so insofar as bravery and camaraderie are objective goods.

Again, however, we are left with the harder cases of adaptive preferences involving domination, injustice, or deprivation. But even here we have the possibility of transformation in which (say) domination or injustice allows a person access to something on a list of objective goods that the person would otherwise lack. For instance, if a child bride was unwillingly married to a dominating older man, but her marriage eventually functioned as a transformative experience that caused her to love him genuinely, then her preference for the dominating marriage could generate some well-being because of the access to love, intimacy, and happiness it provides.

Admittedly, that dominating marriage might provide those objective goods despite rather than because of its dominating nature, and our bride might be prudentially better off if she could access them in a different, nondominating relationship. But the previous discussion of reduction bases applies equally here. If a transformation is psychologically available to our child bride whereby she could become transformed into someone who valued a different and nondominating marriage instead, then she would indeed do better according to our objective list to undergo that transformation. But the transformation caused by her actual dominating marriage might have been immutable, such that new transformations are *not* psychologically available to her.[2] If so, then she would therefore not be capable of accessing many objective goods the nondominating marriage would involve even if she entered into it. Remember that our child bride really does (*ex hypothesi*) get those goods from *this particular* dominating marriage, so it really is her preference for *this dominating marriage* that makes them available to her, and *this particular* dominating marriage therefore really does benefit her even if it also harms her by making other objective goods inaccessible.[3] We might wish that other transformations were available to her, but insofar as they are not, her dominating marriage really does allow our actual child bride to access more objective goods than any other marriage could. So even on objective list accounts, adaptive preferences involving injustice, domination, or deprivation can genuinely prudentially benefit us.

2 Note that for my argument to work, it need not be the case that all or even most people in the child bride's position would be psychologically incapable of this transformation. It need only be the case that it *could* plausibly be true for some. And indeed, even if that failed, we can imagine a nearby case in which the values involved have become so fundamental to her sense of self (or are closely enough tied to other values that are fundamental) that she would be unwilling ever to change her circumstances even when presented with robust opportunities to do so. My argument succeeds just as long as there are *some* adaptive preferences that will be nonvoidable in these ways under these or similar circumstances.

3 While the particular objects that contribute to her well-being are clearly changed through this transformation, note that whether, all things considered, she is better or worse off will depend—especially for the strict objective list theorist—on which objects are included on the objective list and on what their relative weighting is.

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

4. Prudential Good of Pretransformation Selves

Even if a broad range of commonly defended theories of well-being agree that the objects of a person's adaptive preferences can become better than the alternatives for her posttransformation self, do we then have no prudential reason to object to adaptive preferences? Certainly not. Adaptive preferences do not require only retrospective evaluation, after they have been firmly developed. If we take the concept seriously, then preferences must be things that at least sometimes adapt to circumstances over time, rather than being innate, timeless parts of our characters. And if so, then there must be a time before the formation of adaptive preferences, when we could have formed different preferences instead. It is this possibility of forming alternative preferences that allows us to object to adaptive preferences on well-being grounds and to do so on behalf of persons' pretransformation selves.

To illustrate, consider the example of a teenage mother, borrowed from Elizabeth Harman (2009). Having a child will be a transformative experience for teenagers as well as adults. If our young mother—let's call her Rebecca—keeps her unplanned child at age fourteen, she will likely 'raise a child whom she will love dearly. She will love him and be glad that he exists; she will not wish that she had waited to conceive later in life' (Harman 2009: 181). For the sake of argument, let us grant both that this is true of Rebecca and that these feelings are sufficient for her well-being to be retrospectively best served by having and loving *this particular* child at *this particular* time in her life. But even if raising her child is now the prudentially best option for actual Rebecca (let us call her Rebecca_1 for clarity's sake), her life will also be made extremely hard by having a child at fourteen. And there will be another version of Rebecca—Rebecca_2—in a very nearby possible world who waited to become pregnant until she was older, more settled, and more financially stable. Rebecca_2 will not face many of those hardships, and so in that important way she will be better off. Furthermore, Rebecca_2 will not be *worse*-off for lacking the love of the child she would have conceived at fourteen. Instead, Rebecca_2's child will be equally wonderful and equally loved, and her more stable situation will leave her more resources and time to nurture her relationship with that child. Additionally, she will have spent her young adulthood in valuable pursuits that would have been outside of her reach had she been raising a child. So even if Rebecca_1 really is prudentially benefitted by her actual child once it is born, there is an important sense in which Rebecca_2 is nevertheless *better* off for having waited.

Of course, the pregnancy case is in some ways unique, since it involves the introduction of a new and beloved person. Furthermore, the pregnancy case involves an unusually immediate and focused transformation, and many transformations may be slower and more piecemeal. But if we grant for the sake of argument that all transformative experiences lead to complete transformation, then the structure of these other experiences will fit that of the pregnancy case (for discussion of related issues see Terlazzo 2017). I begin at any point in time with a fork in the road, where different paths will lead to different transformative experiences. Different transformative experiences will lead to different possible versions of myself, each of whom will face new forks in the road, with new attendant transformative experiences. At each point in time, we can ask about my

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

prudential interest. Since transformative experiences can change the objects that contribute to my well-being, the different possible versions of myself at any point in time will generally derive the most prudential benefit from different objects. And since adaptive preferences can become the objects that benefit us most through transformative experience, one possible version of myself might be most benefitted by the object of an adaptive preference, while another possible version might be most benefitted by the object of a nonadaptive preference. It is in this sense that, posttransformation, we lack prudential reason to object to adaptive preferences.

However, our Rebeccas illustrate a different sense in which we retain prudential reason to object to worrisome adaptive preferences: it is the reason to prefer that premotherhood Rebecca becomes $Rebecca_2$ rather than $Rebecca_1$.

Whether we can vindicate this claim will once again depend on which theory of well-being we are using. Above, I showed that across a range of theories of well-being, adaptive preferences can become prudentially best for us once we have developed them. In this section, I show on the same range of theories how the preferences discussed above can remain prospectively prudentially objectionable. Given space constraints, I leave it to the reader to translate that explanation for Rebecca's case.

We begin again with simple hedonist and desire satisfaction theories. In addition to its substance, each theory must also explain a life's *degree* of prudential goodness. Hedonists do so by calculating a life's ratio of pleasures to pains, where stronger pleasures and pains count more than weaker ones. Desire satisfaction theorists do so by calculating the ratio of satisfied to frustrated desires, where stronger desires likewise count more than weaker ones. To explain the prospective badness of the dominated housewife's preference, then, they can note that she could initially have undergone a different transformative experience that would have caused her to experience more (and stronger) pleasures and fewer pains or to have more (and stronger) desires satisfied and fewer frustrated. Insofar as her adaptive transformation led her actual posttransformation self to experience more pain and less pleasure overall than her hypothetical counterpart (or to have more desires frustrated and fewer satisfied), it was bad for her pretransformation self to undergo the transformation.

Sumner's authentic life satisfaction account can explain the prospective badness of adaptive preferences by noting that since only authentic—and therefore autonomous—endorsement contributes to well-being, insofar as many adaptive preferences are formed and maintained nonautonomously, any endorsement that they involve will make no well-being contribution whatsoever to a person's life. When and if they become autonomously endorsed (as with our sheltered student), they will begin to make a well-being contribution, but until they do, any endorsement they involve will be prudentially inert. If our student instead initially developed autonomous and nonadaptive preferences, then all endorsement over the course of his life would count toward his well-being.[4] Thus, insofar

4 Cases like Rebecca's are more difficult, since her desire to have her baby may well have been autonomous from an early point. Whether this is so, however, will depend on the details of our account of autonomy, which Sumner does not provide us. Since his account is a species of hedonism, however, we may still be able to say

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

as his transformative experiences could be characterized by autonomy from the beginning, they would be prudentially better for him.

Railton's informed-desire account can again make sense of the prospective prudential badness of the dominated housewife's adaptive preference by appealing to her reduction basis. Our idealized selves would recognize that before we underwent adaptive transformations, a variety of transformative experiences were compatible with our reduction bases. Among the things that our idealized selves would care about would surely be our happiness. Since a nondominating marriage would provide many more opportunities for prioritizing and pursuing happiness, the idealized housewife would not want our initial housewife to undergo the transformative experience of developing the adaptive preference for her dominating marriage.[5]

Finally, the objective list theory can explain quite briefly why worrisome adaptive preferences are prospectively prudentially bad for us. It is prudentially better to have access to more of the objective values on the list and to greater degrees. Although the child bride's dominating marriage gives her access to some objective goods, a nondominating marriage that she can appreciate will likely give her access to many more. The adaptive transformation that gives her access to only a few goods, then, will be prudentially worse for her than (say) the transformation involved in an egalitarian marriage, which gives her access to many more.

Across a range of the most broadly defended theories of well-being, then, we can and must judge the well-being contributions of adaptive preferences from two different perspectives. From the perspective of what benefits an actual posttransformation agent in the moment, adaptive preferences might be prudentially best. But from the perspective of how an agent's life should go over time, undergoing a transformative experience that leads her to *develop* the same adaptive preference might still be bad for her pretransformation self, and having the adaptive preference will *in that sense* be prudentially worse for her. Some adaptive preferences, then, will be both prudentially good *and* prudentially bad for us if we evaluate them—as we can and should—from two different temporal perspectives. But we can make this judgment only if we recognize the role that transformative experience plays in determining our prudential good.

5. Conclusion

I have argued that the same adaptive preference can be prudentially objectionable from one temporal perspective, but prudentially unobjectionable from another. That is, after we have been adaptively transformed, the objects of our adaptive preferences can be better for our actual selves than the objects of their nonadaptive

that while Rebecca_1 autonomously endorses her life, all things considered, the endorsement will in some sense be weaker, given all of the frustrations she experiences.

5 Of course, if the dominating marriage is her unavoidable future, then her idealized self may want her to want it in order to avoid frustration. But like us, her idealized self should be alive to the distinction between what she should do given the options she has and what options it would be best for her to have holding fixed only who she is or could become. I thank an anonymous reviewer for pushing me to clarify this point.

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press

alternatives. However, the same adaptive preferences can remain prudentially objectionable insofar as we could have initially undergone different, nonadaptive transformations in the first place.

One might argue that only the retrospective facts matter. While we might regret adaptive transformations already undergone, our responsibilities are to persons as they are, and we should therefore aim to benefit them as they come to us. I, however, reject this position. Until they are dead, persons always have further transformations ahead of them. Given my argument, it is true that we can have prudential reason to treat persons with already developed worrisome adaptive preferences in accordance with those preferences–but as long as those same persons might develop other worrisome adaptive preferences in the future, we can have equally strong prudential reasons to keep them from doing so in the first place.

ROSA TERLAZZO
KANSAS STATE UNIVERSITY
rterlazzo@ksu.edu

References

Bruckner, Donald. (2009) 'In Defensive of Adaptive Preferences'. *Philosophical Studies*, 142, 307–24.

Dorsey, Dale. (2013) 'Adaptation, Autonomy, and Authority'. In Juha Räikkä and Jukka Varelius (eds.), *Adaptation and Autonomy: Adaptive Preferences in Enhancing and Ending Life* (Heidelberg: Springer), 27–47.

Goodin, Robert E. (2012) *On Settling*. Princeton, NJ: Princeton University Press.

Harman, Elizabeth. (2009) '"I'll be Glad I Did It": Reasoning and the Significance of Future Desires'. *Philosophical Perspectives*, 23, 177–99.

Hurka, Thomas. (1993) *Perfectionism*. Oxford: Oxford University Press.

Kagan, Shelly. (2009) 'Wellbeing as Enjoying the Good'. *Philosophical Perspectives*, 23, 253–72.

Khader, Serene J. (2011) *Adaptive Preferences and Women's Empowerment*. Oxford: Oxford University Press.

Nussbaum, Martha C. (1992) 'Human Functioning and Social Justice: In Defense of Aristotelian Essentialism'. *Political Theory*, 20, 202–46.

Nussbaum, Martha C. (2001) *Women and Human Development: The Capabilities Approach*. Cambridge: Cambridge University Press.

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

Railton, Peter. (2003) *Facts, Values, and Norms*. Cambridge: Cambridge University Press.

Sumner, L. W. (1996) *Welfare, Happiness and Ethics*. Oxford: Oxford University Press.

Superson, Anita. (2005) 'Deformed Desires and Informed Desire Tests'. *Hypatia*, 20, 109–26.

Terlazzo, Rosa. (2016) 'Conceptualizing Adaptive Preferences Respectfully: An Indirectly Substantive Account'. *Journal of Political Philosophy*, 24, 206–26.

Terlazzo, Rosa. (2017) 'When is Non-Ideal Theory Too Ideal?': Adaptive Preferences, Children, and Ideal Theory'. In Kevin Vallier and Michael Weber (eds.), *Political Utopias: Contemporary Debates* (New York: Oxford University Press), 233–52.

Ullmann-Margalit, Edna. (2006) 'Big Decisions: Opting, Converting, Drifting'. *Royal Institute of Philosophy Supplement*, 58, 157–74.

https://doi.org/10.1017/apa.2018.1 Published online by Cambridge University Press