



# Adaptive Preferences Are a Red Herring

**ABSTRACT:** *Current literature in moral and political philosophy is rife with discussion of adaptive preferences. This is no accident: while preferences are generally thought to play an important role in a number of normative domains (including morality, the personal good, and political justice), adaptive preferences seem exceptions to this general rule—they seem problematic in a way that preference-respecting theories of these domains cannot adequately capture. Thus, adaptive preferences are often taken to be theoretically explanatory: a reason for adjusting our theories of the relevant normative domains. However, as I shall argue here, the relentless focus on the phenomenon of preference adaptation is a mistake. While I do not take a stand on whether typical examples of adaptive preferences are or are not problematic, I argue here that if they are problematic, it cannot be because they are adaptive.*

**KEYWORDS:** well-being, Political philosophy, justice, morality, adaptive preferences

In Alfred Hitchcock's classic film noir *Psycho*, we are introduced to Marion Crane (played by Janet Leigh) who has taken a lover in dire financial straits. Early in the film Marion is asked by her employer to deliver \$40,000 to the bank in an effort to complete a real estate transaction. Instead, she absconds with the money and eventually ends up at the Bates Motel.

Marion's theft of her boss's bankroll is a red herring: a plot device intended to distract the viewer from the sinister forces that will eventually be revealed in the person of Norman Bates (and his mother!). Red herrings lead us astray from the main action; they deceive and cover up. Now, in murder mysteries red herrings are part of the fun. But in philosophical contexts red herrings of this sort are apt to lead us to develop theories that are less, rather than more, explanatory, inaccurate rather than accurate, regarding the problems at hand.

In this paper, I argue that the relentless focus in moral and political philosophy in recent years on the phenomenon of *adaptive preferences* succumbs to red herring-itis.<sup>1</sup> My argument will take the following form. The focus on adaptive preferences begins with a kind of intuition: adaptive preferences are somehow problematic in one way or another. But attempts to explain why adaptation is

<sup>1</sup> Notice that I am not the first to claim that adaptive preferences should be regarded as red herrings. To my knowledge, the first to do so is Baber (2007). However, Baber's claim is that adaptive preferences strictly speaking do not really exist as cited by Sen and Nussbaum; see below. My argument, instead, is that though adaptive preferences may (and, plausibly, do) exist, the phenomenon of preference adaptation should have no particular explanatory role when it comes to moral theory choice.



problematic must either fail or succeed but at a cost of subverting the initial intuition about preference adaptation. But if this is correct, the phenomenon of specifically *adaptive* preferences simply drops out of any plausible account of moral, political, or axiological theory. And hence, like the \$40,000 that ends up buried with Marion's car, adaptive preferences are a red herring.

My intention here is not to defend a substantive theory of practical rationality or the personal good according to which common examples of adaptive preferences are, in fact, not problematic (for such arguments see, e.g., Bruckner [2009]; Smith [2013]). Rather, my claim is more general. I claim that the explanation for the problematic nature of such preferences, assuming they are problematic, cannot be compatible with the normative significance of adaptation. The category of adaptive preferences should cease to play an interesting role in our theorizing, whether or not typical instances of adaptive preferences are problematic from the point of view of moral, political, or axiological theory.

## 1. Adaptive Preferences and a Skeptical Challenge

The topic of adaptive preferences has occupied much recent moral and political philosophy. Representing a broad consensus, Serene Khader writes that adaptive preferences are 'morally distinct . . . a category of preferences that are particularly worthy of questioning by public institutions' (Khader 2011: 47).

The argument of this paper is meant to show that treating adaptive preferences as a special, morally significant category is a mistake. But before I begin, a sensible question requires answer: what are adaptive preferences?

Of course, it is difficult to provide a unifying account of the nature of adaptive preferences, especially given the fact that much of the philosophical attention that this phenomenon has drawn has been on the basis of a few provocative examples. Sen, for instance, writes:

Our desires and pleasure-taking abilities adjust to circumstances, especially to make life bearable in adverse situations. The utility calculus can be deeply unfair to those who are persistently deprived: for example, the usual underdogs in stratified societies, perennially oppressed minorities in intolerant communities. . . . The deprived people tend to come to terms with their deprivation because of the sheer necessity of survival, and they may, as a result, lack the courage to demand any radical change, and may even adjust their desires and expectations to what they unambitiously see as feasible. (Sen 1999: 62–63)

Nussbaum suggests a particular instantiation of Sen's more general account:

Consider Jayamma. . . . When women were paid less for heavier work at the brick kiln and denied chance for promotion, Jayamma didn't complain or protest. She knew that this was how things were and would

be. . . . When her husband took his earnings and spent them on himself in somewhat unthrifty ways, leaving Jayamma to support the children financially through her labor, as well as doing all of the housework, this didn't strike her as wrong or bad, it was just the way things were, and she didn't waste time yearning for another way. . . . Jayamma seemed to lack not only the concept of herself as a person with rights that could be violated, but also the sense that what was happening to her was a wrong. (Nussbaum 2000: 113)

While these examples are evocative, it is worth noting the vast diversity of examples of adaptive preferences and the accompanying challenges in clearly identifying the concept. Khader helpfully catalogs this diversity in canvassing a number of important cases. For instance, in such cases the processes with which adaptive preferences are formed differs; some processes are conscious, some unconscious, some are 'described as adaptive without any discussion of whether they were consciously formed'. In addition, Khader notes that some examples show preferences 'that seem to exist because of a lack of options . . . and preferences that seem to be formed because of a lack of awareness'. In addition, diversity exists concerning whether these preferences undermine 'people's entire senses of self' or do not (Khader 2011: 10).

Despite this diversity, however, a key feature seems to be shared by typical examples of adaptive preferences. Adaptive preferences are shaped (by whatever process and to whatever extent) by facts of, or perceptions of, availability or possibility. (Khader seems to suggest that some examples merely show a 'lack of awareness'. But a key factor here is that the example she discusses in this domain is explicitly a lack of awareness of possibilities [Khader 2011: 9].)

This feature seems to form the typical understanding—which I shall follow here—of adaptive preferences. Rosa Terlazzo writes: 'The idea behind the concept of adaptive preferences is roughly this: that when persons have limited option sets, they can come to prefer things within those sets that they would not prefer otherwise' (Terlazzo 2015: 179). According to M. Rickard, 'with adaptation, the preference for one option over another varies with the perceived availability of each option' (Rickard 1995: 279). For Elster, individuals lack adaptive preferences if it is the case that the facts of feasibility have no general effect on their preference structure (Elster, 1982: 229.) Bruckner holds that 'one type of adaptive preference is a preference that changes in response to the contraction of the set of options that are feasible for the agent, that is, capable of being attained, (Bruckner 2009: 307). Elizabeth Barnes holds that adaptive preferences are 'formed toward something sub-optimal in light of a severely diminished set of options' (Barnes 2009: 5).<sup>2</sup> While these accounts are clearly nonidentical, they share the general thought that if my preference for  $\phi$  is adaptive, the fact that I prefer  $\phi$  must be explained, at least to some extent, by facts of the unavailability or perception of the unavailability of

<sup>2</sup>The 'suboptimal' qualifier here stands in some need of elucidation. As I argue below, it should not be understood as *objectively* suboptimal. Rather, it is best understood as something like: 'suboptimal in light of the preferences one would have had absent option-constriction'.

some alternative to  $\phi$ . This is a pretty wide understanding of adaptive preferences. But I resist making this account any more specific insofar as my argument is meant to be broadly applicable to many different proposals. Nothing in my argument will hinge, for instance, on the more precise nature of the underlying preferences or the various mechanisms by which they are generated given facts of unavailability. Of course, some may define adaptive preferences differently, but it is clear that identifying adaptive preferences as those that are shaped by facts of availability has broad resonance in the literature on this phenomenon.

A second feature of these examples is worth noting. Clearly, adaptive preferences seem to be problematic. Immediately when confronted with cases like that of Jayamma, we are tempted to think that we should not take such preferences seriously—as Sen suggests, to do so would be perhaps unfair or wrong. When individuals adapt to their circumstances or limitations, we tend to think that such adaptation should not determine their good or what justice demands. I am going to call this reaction to such cases the *adaptive preference intuition*: the fact that a preference is adaptive is reason to believe that it is problematic.

It may do to say a bit more about what it means to call adaptive preferences problematic. Typically, adaptive preferences are thought problematic insofar as they appear to be exceptions to the generally plausible explanatory role preferences play or are thought to play in particular domains of inquiry. For any domain (such as, e.g., political justice), adaptive preferences are problematic to the extent that while it is generally true that preferences (the ‘favored’ preferences, whatever those are) will play a role in that domain, adaptive preferences cannot play the same role. The adaptive preference intuition, then, seems to indicate that the phenomenon of preference adaptation is good evidence that the relevant preferences are not favored for preference-sensitive domains.

Adaptive preferences are thought especially problematic in three typical domains. First, in the domain of well-being or personal good it seems generally plausible to say that the fact that I prefer  $\phi$  to  $\psi$  is at least *some* evidence that  $\phi$  is better for me than  $\psi$ , however else one’s theory of welfare is constructed. However, adaptive preferences are thought to lack the requisite authority to determine the objects, events, and states of affairs that are (even *pro tanto*) intrinsically good for those persons whose preferences they are. Second, in the domain of morality many believe that proper moral consideration of a person will have at least something to do with respecting this person’s preferences. But that my preference for  $\phi$  is *adaptive* would seem to be a defeater of this effect: adaptive preferences, unlike other preferences, cannot justify (from the moral point of view) treating me in some way that would otherwise be ruled out. Third, in the domain of political justice, just institutions are typically thought to take seriously the preferences of citizens in any number of ways, for example, through the operation of democratic voting and the rejection of paternalistic intervention. However, if such preferences are adaptive, there may be good reason not to take seriously that preference as a matter of justice. For instance, despite the fact that their preferences are directed toward their poor circumstances, Nussbaum writes that nevertheless ‘we have good reasons, for example, to support public investment in female literacy, even in the absence of young girls’ demand for such programs’ (Nussbaum 1999: 151).

At this point, however, though the adaptive preference intuition may be quite robust, there is a skeptical challenge that seems to me worth answering. To say that adaptive preferences are problematic is to say that the causal origin of preferences gives these preferences a particular role in theory construction—in other words, in any acceptable first-order theory, that a preference is *caused* in a particular way seems to rule that preference out (or at least seems to me a reason to rule that preference out) of consideration. But why? After all, many of our preferences are in one way or another shaped by our circumstances (Bruckner 2009: 316–17). I may grow up in Kansas City and develop, as a result, a preference to see the Royals win the World Series. I might have a parent who is an academic and develop a preference to take on a career that involves substantial time on college campuses. If these preferences count as favored—preferences that are specifically a result of the seemingly arbitrary circumstances in which we are placed—why should preferences that are specifically a result of limitations be not favored? Why should, in other words, adaptation be at all interesting in understanding the normative force of preferences? This line of inquiry plausibly gives rise to a

*Skeptical Challenge:* although there may be many reasons for a preference to be nonfavored, that this preference is adaptive is not one of them.

Insofar as the skeptical challenge is, in essence, the denial of the adaptive preference intuition—and seems to be buttressed by a plausible (or, at least, plausible enough) rationale—defeating the skeptical challenge requires that we actually vindicate the adaptive preference intuition by explaining why preferences that are shaped by limitations in particular should have no normative role.

## 2. What Makes Adaptive Preferences Problematic?

What would such an explanation look like? Presumably, there should be some feature of adaptive preferences that would explain why adaptive preferences ought to be treated as nonfavored. Furthermore, to respond to the skeptical challenge this feature must pass two important tests. First, it should explain why adaptive preferences should be regarded as problematic in light of the fact that many preferences arise specifically as a result of contingent, morally arbitrary circumstances (like one's parents' profession or one's geographical location). I am going to call this the *explanatory test*. Second, and crucially, it must be the case that the relevant property does not render the fact that the preferences in question are adaptive a red herring. In other words, it must be compatible with the adaptive preference intuition—that preferences are adaptive is reason to believe that they are problematic. Call this, for lack of a better term, the *red herring test*.

Here is one way a particular explanation could fail the red herring test. Take an analogy. Imagine that your experience with cars built by Chevrolet is that such cars do not run well. For instance, whenever you are assigned one from a rental car company, it turns out worse than you expect; your friends' Chevy cars always

seem to be breaking down, and so on. Call your general sense of the shoddiness of Chevy cars the *Chevrolet intuition*. But this calls for explanation: why should it be that Chevy cars do not make the grade? Why should Chevrolets be, as it were, ‘car-problematic’? Imagine that we discover that the cars you have experience with have engines that are built in a particular plant that has a history of building engines poorly. Now imagine that having an engine built in this plant really does make cars that have such engines run poorly and that this does, in fact, explain why cars built there are car-problematic (in other words, the argument passes the analogical explanatory test). However, imagine that it is also true that not all Chevy cars have engines built in this plant and also it is not only Chevrolet that uses such engines, but many other car manufacturers do as well. If this were the case, though we may have explained why some cars you have experience with run poorly, we will also have determined that the *fact* that some car is a Chevy is no reason to believe that it is car-problematic. In coming up—stretching the metaphor here, apologies—with a proper theory of the best cars (or, at least, of which cars do and do not run well), the fact that some car is a Chevy is a red herring. And so it is, or so I shall argue, with adaptive preferences.

In what follows, I am going to examine a number of potential accounts of what makes adaptive preferences problematic. I will not argue for any one of them in comparison to the others. Instead, I will assume for the purposes of argument that each (save one, see [section 2.4](#)) passes the explanatory test. But the question will be whether, in so doing, the relevant explanation can pass the red herring test. I argue that the answer is ‘no’.

### 2.1. Inconsistency with Objective Measures

Consider paradigmatic examples of problematic adaptive preferences from, for example, Sen. In each case, the preferences of these individuals are shaped by their limitations. But another thing is certainly noticeable about them. Their preferences seem to point away from what is good, right, or just, and instead point toward something bad, wrong, unjust. And if this is correct, it could be that the problematic feature of preference adaptation is that it yields adaptation toward objectively worse states: for the domain of well-being, the worse rather than better for the person; for the domain of morality, the morally worse; for political justice, the comparatively unjust.

Indeed, this proposal has been suggested by a number of theorists of adaptive preferences. In attempting to offer a version of consequentialism that can avoid adaptive preferences, M. Rickard proposes an ‘objective’ consequentialism, according to which ‘the value of states could be plausibly interpreted in preference-independent terms’ (Rickard 1995: 296). A similar view is proposed by Jennifer Hawkins (Hawkins 2008: 167). Famously, Nussbaum suggests that adaptive preferences are problematic when they are incompatible with the maintenance of her objective list of ten basic capabilities (Nussbaum 2000).

However, if the problem to be found with adaptive preferences is that they are inconsistent with objective measures of the domain in question, it is difficult to see what specific role the fact that these preferences are adaptive should play

in the theory ultimately developed. To put the question bluntly, if inconsistency with objective measures explains why a preference is problematic, why should it matter that this preference is adaptive? It would seem that if we have good reason to treat preferences that are inconsistent with objective measures as nonfavored, then the fact of adaptation seems entirely beside the point. After all, it is certainly possible for me to develop a nonadaptive preference that is inconsistent with objective measures of a particular domain. I could, nonadaptively, come to prefer some state that is not good for me or that is morally illegitimate or that is inherently unjust. Furthermore, there is nothing that rules out the possibility of adaptively preferring a state that is not inconsistent with objective measures. Imagine, for instance, that I long desired to be a world-class trumpet player, but that my lips simply could not produce the embouchure required. But I come to adapt my preferences given my lack of options toward being a world-class trombone player (for which I do have some talent). It would be a bizarre suggestion that being a world-class trombone player is inconsistent, even relative to being a world-class trumpet player, with objective measures of any of the relevant domains. But this preference is clearly adaptive. And if that is correct, inconsistency with objective measures cannot justify any theoretical role for adaptive preferences in particular.

And thus if we explain the adaptive preference intuition by holding that preferences that are inconsistent with plausible objective measures are problematic (if, in other words, the current proposal passes the explanatory test), the fact that a preference is adaptive appears to be no evidence that this preference is problematic. This explanation therefore fails the red herring test.

## 2.2. Autonomy

A second (popular!) possibility is that adaptive preferences, unlike other preferences, are nonautonomous. This proposal is floated or at least discussed by Elster (1982), Sumner (1996), Colburn (2011), Terlazzo (2015), Taylor (2009), and many others. This view (purportedly) can distinguish adaptive preferences from those preferences that maintain explanatory power relative to a given domain independently of any consilience with an objective set of measures for the domain in question. Thus the problem with Jayamma's preferences, for example, is not that they are preferences for an objectively bad state of affairs, but rather that they are nonautonomous.

However, it is immediately unclear what is meant when we are told that adaptive preferences are nonautonomous. Insofar as autonomy appears to be a concept in need of a conception, one must be given an account of the autonomy of preferences before one can hope to explain the problematic feature of specifically adaptive preferences in those terms (see Terlazzo 2015: 183–84).

For the purposes of this paper, I shall focus on two criteria that seem to have some currency with those who would label adaptive preferences nonautonomous. Call the first the mechanistic approach. On this view, autonomous preferences (or their opposite) are identified by their specific formation mechanisms. Elster writes:

If this definition of freedom is to be of real value, we need a definition or a criterion of autonomous wants. This I cannot provide. I can enumerate a large number of mechanisms that shape our wants in a non-autonomous way, but I cannot say with any confidence whatsoever that the wants that are not shaped in any of these ways are *ipso facto* autonomous. (Elster 1982: 228)<sup>3</sup>

Sumner says something similar:

It appears, therefore, that neither of the currently dominant theories about the nature of autonomy is self-sufficient. Theories which emphasize the importance of identification must provide some reassurance that the agent's values or standards have themselves been autonomously adopted, while theories which look backward to this formation process must require preservation of the capacity for critical distance and reflective assessment. However the details of a fully adequate view are worked out in the end, the implications for our theory of welfare are clear. Self-assessments of happiness or life satisfaction are suspect . . . when there is good reason to suspect that they have been influenced by autonomy-subverting mechanisms of social conditioning, such as indoctrination, programming, brainwashing, role-scripting, and the like. (Sumner 1996: 170–71)

However, the problem here should be obvious. If preferences are suspect when they are produced by nonautonomous mechanisms (however clearly delineated), there is very little of interest in the fact that they are adaptive, that is, produced as a result of a person's limited possibilities. After all, my preferences could be a result of 'role-scripting' (or any other suspect mechanism) even if my resulting preferences have nothing to do with my limitations. In addition, I might decide, after calm, rational, fully informed deliberation that my preferences are simply unsatisfiable and choose to undergo a process of preference revision given this fact. This mechanism is clearly autonomous but nevertheless produces adaptive preferences. Thus, on the mechanistic approach the fact that some preference is adaptive simply drops out—the skeptical challenge is unsolved.

One response to this claim is worth considering here. Elster and others insist that there is a big difference between preferences that are adaptive and preferences that are a result of 'deliberate character planning'. For Elster, adaptive preferences take place 'behind my back', as a result of a 'drive' rather than a second-order desire. (Colburn [2011: 67–70] also suggests that adaptive preferences must operate 'covertly'.) According to Elster, for instance, there is no genuine problem of preferences that are consciously changed to fit one's circumstances. 'If I consciously shape myself so as only to want what I can get, I can attain full satisfaction of an

<sup>3</sup> Elster also offers some additional evidence that one's preferences are nonautonomous, viz., that there is preference reversal among options in one's feasible set (see Elster 1982: 229). At best, however, this evidence is imperfect, as Elster fully admits.



autonomous want structure, and this can with more justification be called freedom, in the Stoic or spinozistic sense' (Elster 1982: 227). However, it is important to ask the right question here. Elster suggests that deliberate character planning results in preferences that are not problematic—and he does this by simply agreeing with my suggestion that, for example, rational deliberation to alter my preferences is autonomous. But whether or not such preferences are problematic, the relevant question for current purposes is whether such preferences are adaptive. And it appears to me that there is no plausible reason to claim that preferences formed by deliberate character planning are not adaptive.<sup>4</sup> Of course, one might simply define adaptive preferences as specifically an 'unconscious altering of our preferences in light of the options we have available' (Colburn 2011: 52). But, first, I suggest that this is an unwarranted limitation on the concept of preference adaptation, as already argued. But even if we allow that adaptive preferences must be unconscious, it is of very little help for the mechanistic approach. I can, in nondeliberate (i.e., unconscious) but nonetheless anodyne ways, adapt my preferences to my circumstances, just as I might nondeliberately (unconsciously) but in a nonetheless anodyne way find myself a fan of the Kansas City Royals merely as a result of my geographic location. (See, for instance, the diabetes case below.) Such a preference, though covert, would not fail any plausible mechanistic account of autonomy though it would clearly produce preference adaptation.

However, there is a second criterion for the autonomy of preferences. One might also identify the autonomy of preferences not by the specific formation mechanisms involved, but rather given some fact about the preference *now*, in particular, whether or not the right sort of relation holds between the preference and the agent whose preference it is. Call this a 'relational' account of the autonomy of preferences. What sort of relation? One could construe it in many ways; one might say that the right sort of relation obtains between a preference and the person in question when this person endorses or prefers to maintain this preference (Colburn 2011: 68; see also Frankfurt 1971; Lewis 1999). On this view, it is sufficient to maintain an autonomous preference if that preference is the object of a second-order pro-attitude on the part of the agent in question. Alternatively, it could be that a preference is autonomous if that preference coheres with, or withstands the scrutiny of, other first- and second-order evaluative attitudes maintained by the relevant person. Though there may be other ways to understand the nature of the 'right' relation, broadly speaking this account of the autonomy of preferences holds that a preference is autonomous if that preference reflects or coheres with the 'real self', which is in some way determined by a favored subset of one's own valuing attitudes.

But as with a mechanistic account, a relational approach to autonomy fails the red herring test. If one accepts a relational account, there is surely no guarantee that

<sup>4</sup> Elster writes that while deliberate character formation can result in a person's preferences being identical to their opportunities, adaptive preferences tend to 'overshoot, resulting in excessive rather than proper meekness' (Elster 1982: 224). Second, adaptive preferences typically downgrade unavailable options rather than pumping up the available ones, like deliberate character formation (Elster 1982: 224). None of these differences, however, suggest that deliberate character planning is not genuinely adaptive.

all nonautonomous preferences will be adaptive—I can develop a nonautonomous preference as a result of addiction, for example—such a preference may well lack the ‘right relation’ to my genuine valuing states (Lewis 1999). And while addictive preferences may very well be problematic from the perspective of morality or the personal good, addiction is certainly not adaptation. Furthermore, it is not the case that all adaptive preferences will lack the ‘right relation’. Imagine that early in life I was diagnosed with type-1 diabetes. To cope with this devastating diagnosis, however, I began to look at myself as *essentially* diabetic: I saw this disease as crucial to my own identity and understanding of myself. This preference was adaptive: adaptive to the conditions of my illness and the limitations I faced. (Furthermore, it is worth noting that this example can be described as operating covertly, as suggested by Colburn) One may question the extent to which this preference was autonomous at the time—perhaps this preference was inconsistent or incoherent with other things I valued. But even if this preference had been nonautonomous at the time at which it was formed, there is no reason to believe that, over time, this preference could not *become* part of the ‘real me’ (Sher 1997: ch. 3).

A reasonable reply to this is that when a preference reflects the ‘real me’ does this not entail that it is *not* adaptive? After all, it would appear not explained by the limited option sets I maintain. But this is not accurate. It could be that the overall set of evaluative attitudes I maintain would change radically if my option sets were to open. But given that my option sets are constrained, the preference itself maintains whatever relevant relation one specifies to the other evaluative states one maintains. But it would be strange to say that just because one’s entire (including second-order) evaluative set is adaptive, no particular part of the set is adaptive. One might be tempted at this point to hold that, so far as this preference has really become part of the ‘real me’, it is not genuinely problematic. Perhaps this is plausible (Dorsey 2012). But this conclusion simply reiterates the extent to which the fact that preferences are adaptive is neither here nor there.

### 2.3. Status-Quo Bias

If the explanation of the problematic nature of adaptive preferences is that they are inconsistent with objective measures or are nonautonomous, adaptive preferences are red herrings. But there are other possible explanations. To see the next, reflect on what seems to be so problematic about adaptive preferences from, for instance, a moral perspective. To suggest that moral concern for a person should be determined by what he or she adaptively prefers would seem to inflate the moral value of the status quo artificially. After all, if the person in question simply prefers the status quo because her options to break out of the status quo are limited and if moral concern for this person involves respecting her preferences, it would seem that the current state of affairs (for this person) is artificially morally favored (or at least has a positive moral status), just because it is the status quo.

In previous work, I have suggested that the real problem with adaptive preferences (at least when it comes to the domain of well-being) is that they violate the following principle (which could trivially be adapted to other domains):

*Independence*: Whether  $\phi$  would be better for  $a$  than  $\psi$  at  $t$  is independent of whether  $\phi$  rather than  $\psi$  is more likely to occur (Dorsey 2010: 539).

A few caveats should be offered here.<sup>5</sup> Of course it is true that when it comes to the promotion of morally valuable or just states of affairs, states of affairs that are better for us, and so forth, whether or not this state of affairs is likely or unlikely to occur can and should play a role in our deliberation. Because states of affairs that are unlikely to occur require greater time and energy to make happen (if they happen at all), it could be that aiming for more realistic options is a good principle of prudential or practical reasoning (Nebel 2015). But this does not tell against Independence. There may be good instrumental reasons to prefer the near to the far—but it should not be the case that the fact that something is more likely to occur by itself renders this state morally or prudentially or politically favored. Instead, Independence is an account of the objective reasons (in the case of morality) or an account of what actual states do or do not make someone better off.

However, if violations of Independence are what renders adaptive preferences problematic, the same problem that plagued previous accounts arises again: that a preference is adaptive is no evidence that it is problematic. To begin with, consider a person who, at least when it comes to the typical set of human activities, is without substantive constraint or limitation. Perhaps, this person is fabulously wealthy, socially favored, talented in all relevant ways, and so on. Of course, this person may have some limitations: she cannot square a circle, cannot live to be 200 years old, etc. But for all intents and purposes, she can do whatever she wants, live in whatever way she wishes. It would seem impossible for this person to have genuinely adaptive preferences, at least when it comes to the activities we are focusing on here. But surely it is the case that this person's preferences can violate Independence. For instance, she might prefer for its own sake a home-cooked meal to a meal at a restaurant. Asked why she prefers this, she might say in all seriousness that the home-cooked meal is more likely to occur than the restaurant meal. After all, everything she must do to procure a restaurant meal makes it slightly less likely that this meal will be procured. Of course, it is not because she faces a limitation: in every meaningful, reasonable sense of 'can', she can go out to a restaurant if she wants. This may seem strange, but it is certainly not unintelligible: she just prefers to eat at home because a home-cooked meal is (marginally) more likely. It may very well be plausible to say that this preference is problematic; after all, it seems quite clearly to violate Independence, and Independence may very well be a plausible principle of the personal good (or morality or political justice). But this is surely

<sup>5</sup> One might object that independence (and its cognates) do not really describe preferences that display a status quo bias, but instead a bias toward the likely; for many reasons it could be that the status quo is less likely to continue to occur than some change in the status quo. While this may be true, it is clear that the relevant bias on display in adaptive preferences is a likelihood bias. But, second, this makes no genuine difference. Even if we tweaked independence to reflect a true status-quo bias, this would be no less problematic. One could simply change the example I offer below to a person who genuinely desires to keep things as they are (where this is read *de dicto*). This would be a quirk, but it could violate the revised versions of independence and would not be adaptive.

not an adaptive preference. Her psychology just features an Independence-violating quirk. Indeed, any set of preferences that track the likelihood of particular states (whether or not such preferences are adaptive) will violate Independence, whether likelihood is preferred (as in the above case) or dispreferred (as in, for instance, a person who prefers lost causes).

Of course, even if not all Independence-violating preferences are adaptive, one might think that given that adaptive preferences are defined as those preferences that develop as a result of one's limitations, they necessarily (perhaps even as a matter of concept) violate Independence. After all, the fact of a limitation just is the fact that some state of affairs is unlikely to occur. And if this is correct, it would appear that a violation of Independence (or its cognates) is sufficient to explain the problematic nature of adaptive preferences even if it also explains the problematic nature of lots of other preferences. On these conditions, one could continue to accept the adaptive preference intuition and simply maintain that adaptive preferences are not the only problematic ones. (Compare: though many other cars are made by the poorly performing factory, because all Chevrolets are made there, *avoid Chevrolets*.)

But this does not help. Notice that Independence, as stated, holds that the relevant feature is the extent to which the particular preferred objects or states are rank-ordered as more or less valuable based on their likelihood. However, as already suggested, it is plausible to say that a preference qualifies as adaptive if it is explained by the *perception* of one's limited option sets. However, it is quite clear that one can perceive that one's options are limited to some set *S*, and hence have one's preferences adapt to *S*, but for it to be the case that one's option set is much wider than *S*, and hence the set to which one has adapted is no more or less likely to occur than others that are beyond the range of one's adaptation. I might, for instance, believe that I am only cut out to be a golfer, not a basketball player, and adapt my preferences to the former when in fact my options were much wider—it was perfectly possible for me to be a basketball player if only I had tried. This preference would not violate Independence, but it is clearly adaptive. One might hold that preferences that are explained only by perceived limited option sets, rather than by actual limited option sets, are not genuinely adaptive. But that is straightforwardly implausible on its face and furthermore an ad hoc attempt to squeeze all adaptive preferences under the heading of Independence.

Much more plausible, instead, is the claim that though Independence may pass the explanatory test, like the two previous explanations, it cannot pass the red herring test.

## 2.4. Adaptation

In a way the argument of the previous three sections might seem a touch on the obvious side. If we are explaining the features of adaptive preferences that render them problematic in a way that is abstracted from the fact that they are adaptive, surely we are going to end up with some explanatory feature that need not bear an essential connection to the fact that preferences are, in fact, adaptive. But we can avoid this by instead claiming that the best explanation is the fact that preferences

are adaptive; they arise given facts about our limited option sets. This is what makes them morally problematic, both problematic when it comes to understanding our own welfare and problematic for the sake of political justice.

But this proposal cannot respond to the skeptical challenge. This challenge is worth answering, and to respond by claiming that the adaptation of preferences is just a brute explanatory feature fails to take it seriously. To resist the skeptical challenge we need to know why specifically adaptive preferences are problematic in comparison to other humdrum preferences that may arise as a result of morally arbitrary factors (such as one's geography or upbringing). (Compare: one would surely reject as nonexplanatory the suggestion that Chevrolets run poorly just because they are Chevrolets. One would expect some explanation about their construction, their fuel efficiency, and so on.) Merely to identify the fact that preferences are adaptive as the feature that renders them problematic does not seem to answer to our interest in coming up with an explanation for the problematic nature of adaptive preferences. We already know these preferences are adaptive. We want to know *why* being adaptive makes them problematic. Thus, the problem here is the reverse of the previous attempts to explain why adaptive preferences are problematic: rather than failing the red herring test on the assumption that it passes the explanatory test, this proposal seems to pass the red herring test only to fail the explanatory test.

## 2.5. A Mixed View—and a Generalization

The final proposal I will consider here is offered by Khader. Rather than suggesting that there is a unified explanation of what makes adaptive preferences problematic, why not offer a conjunctive or mixed view? According to Khader, the problematic nature of adaptive preferences can be explained by the fact that they are not only inconsistent with objective measures (“inconsistent with basic flourishing”), but are also formed as a result of conditions that are unsuitable to basic flourishing (Khader 2011: 42–49). Defending this sort of approach, Khader claims that

if adaptive preferences seem morally distinct, if they seem like a category of preferences that are particularly worthy of questioning by public institutions, it is both because they are flourishing inconsistent and because they seem imposed on their bearers. Remove either feature from a case, and we no longer seem to be left with adaptive preference. (Khader 2011: 47)

However, I find this suggestion puzzling. Put bluntly: why is it part of an explanation of the problematic nature of adaptive preferences for human flourishing that they are caused or formed by conditions that are unsuitable to basic flourishing? Why is it not a fully sufficient explanation that they are inconsistent with basic flourishing? Of course, one might hold that to be formed in a particular way is part of what makes these preferences adaptive. But even if we agreed with this suggestion (which, as already argued, we should not—there is no reason why, for a

preference to be adaptive, it must be caused by conditions inconsistent with basic flourishing), this fact surely adds no additional information concerning why such a preference is problematic. Indeed, so much is implied by the argument of the previous section—given that the fact that a preference is adaptive does not explain why it is problematic, all the explanatory heft, in Khader’s account, must be given by its inconsistency with objective measures, leading to the result that adaptiveness is a red herring. Put another way: insofar as Khader’s view passes the explanatory test, it fails the red herring test.

Khader resists this by presenting a case in which the person chooses a life that is inconsistent with her flourishing but is nevertheless nonadaptive. She writes:

We might imagine a woman who has access to conditions for flourishing across a variety of domains, who reflects upon her decisions, and who decides in spite of it all to spend the rest of her life engaging repeatedly in a sport that is likely to kill her—say bullfighting. . . . This woman seems not to have adaptive preferences. We may have paternalistic desires to prevent her from bullfighting, but we would not think of her preference as adaptive if social conditions did not seem to have unduly influenced this choice. (Khader 2011: 48)

In response, however, I would have to say that it is unclear that this case is one in which the woman’s preferences are inconsistent with her flourishing, even on Khader’s preferred perfectionism. Second, Khader’s claim here is simply that her preferences are not adaptive. And I agree! But this hardly means that if they are not adaptive, they are not problematic.

Of course, we might try to piece together some further view, suggest some other explanation. But at this point I would like to make something of a generalization. There is good reason to believe that we will not find an adequate explanation of the problematic nature of adaptive preferences that does not render them a red herring. To begin, it is plausible to hold that not all adaptive preferences are problematic. And hence there is good reason that none of the proposed explanations so far considered rule out all adaptive preferences: not all adaptive preferences should be ruled out when it comes to preference-dependent theories of welfare, morality, and political justice. Take my preference in favor of being a type-1 diabetic. This preference is, in fact, adaptive: it was formed as a result of my own limitations. (Furthermore, it seems to fail not only the relevant objective measure, but also was formed as a result of conditions unsuitable to my own flourishing, i.e., chronic illness.) But as time has progressed, I have come to identify strongly with it; my psychology has, as it were, changed around this particular preference. Of course, this preference remains adaptive—the limitation is still there, it is this limitation that explains the maintenance of this preference. It is just now the case that I—in part given my limited option sets—very strongly identify with this preference.

Is this preference problematic? First, it is hard to see how it could be problematic from the standpoint of the personal good (see esp. Barnes 2009: 8.) If I deeply identify with being a diabetic, it would be strange to say that the state of affairs

in which I am cured would be, in itself, good for me.<sup>6</sup> And the explanation must surely make reference to the fact that I maintain the relevant preference. We might say that were the option to be cured to open up and were my preferences to change, then to be cured may be a benefit. But, of course, this is not what we are imagining. Under those conditions, at that time, a cure may benefit me. But the question we are asking is whether, given my preferences as they are, my conditions as they are, the cure would be better for me than being a type-1 diabetic. Plausibly, no. For similar reasons, it is hard to see how this preference would fail to be relevant to morality or political justice. Given my preference, there appears to be a reason of justice to take very seriously my interest in being diabetic and, for instance, treat my preference as reason to design social institutions that are accessible to diabetics rather than (as a matter of justice) simply seeking to alleviate the underlying condition itself. (Of course, this is compatible with suggesting that on the whole the best political policy would generally be to cure diabetics; it would simply be that *my* being a diabetic, given my preference, would not be such a reason.<sup>7</sup>)

Of course, one might be squeamish at suggesting that such preferences are not problematic. After all, to allow that they are not so problematic seems to allow in too many forms of adaptive preferences, preferences that we would prefer to expunge. For instance, it might be said that this very phenomenon is what is happening in the case of Jayamma or other individuals who adapt to particularly oppressive conditions. It might be that, over time, the psychology of such individuals changes around the relevant preference, such that they come to identify with their own oppression. And although not the particular form of oppression by choice discussed by Ann Cudd (1994), this sort of 'deep adaptation' may give rise to important forms of oppression by choice: it may be that for some oppressed individuals it is the case that they have been down so long, it looks like up. But surely we would not want to say that *these* preferences generate moral or political reasons (at least—perhaps we may be cautiously willing to admit that these preferences are not problematic when it comes to the domain of well-being).

I think this is quite sensible. But—and this is the second argument—cases like this provide further evidence that adaptation, by itself, is a red herring. There may very well be some subset of adaptive preferences that we wish to rule out. But that subset must be identified by some feature—adaptive preferences that are a result of oppression by choice, for instance. But in making that identification, it is hard to see how preferences that display that feature are problematic *only* when they are adaptive. Presumably, if oppression by choice is a problem from the standpoint of moral or political theory, it is a problem whether or not preferences that lead to oppression by choice are specifically adaptive to that condition.

<sup>6</sup>Notice that this need not rely on a deeply subjectivist theory of well-being; it would be enough for the view simply to require that states that are radically disvalued by the person in question are not comparatively beneficial; this proposal is broadly ecumenical, at least among views for which preferences and their favored or nonfavored status is a relevant issue.

<sup>7</sup>Thanks to Rosa Terlazzo for helpful critical comments here.

One might deny this! One might hold—with Khader—that the particular feature that identifies the relevant subset of adaptive preferences alone cannot explain their problematic nature. The fact that they are adaptive must continue to play a role. But this just brings us back around to the skeptical challenge all over again—why should adaptation play this role? To answer this, one must offer some explanation—and we are back where we started.

Of course, there may be some as-yet-undiscovered explanation that shows, once and for all, that adaptation is really relevant from the point of view of theory. But at this point we are entitled to a bit of inductive generalization. After all, for each answer to the skeptical challenge so far discussed, we have found that adaptation simply drops out as explanatorily interesting. If this is correct, there is good reason to believe that adaptive preferences are not just a red herring according to the various accounts that have so far been canvassed, but the fact that we have yet to find an account that entails that adaptive preferences are not a red herring is an indication that they really *are* a red herring.

### 3. Objection: So What?

So far, I have argued that the fact that preferences are adaptive seems to play no genuine role in explaining why those preferences do or do not capture the facts of individual well-being or generate moral or political reasons.

One might legitimately question why this fact should be worthy of notice for three reasons. First, while it may be the case that even though preference adaptation does not play an ultimate explanatory role in crafting our favored normative theories, it may still be the case that an examination of adaptive preferences leads us to develop the right sort of view. This is compatible with suggesting that many adaptive preferences will play the role of favored preferences and that preferences that display flaws other than adaptation will not. It is just that some forms of adaptation are glaring examples of problems with particular preferences that may extend beyond adaptation (though not necessarily to all forms of preference adaptation). (Compare: we find out that lots and lots of Chevrolets do not run well—and the search for the underlying problem leads us to discover that the factory in question produces shoddy engines.)

And while this may be true as, say, a historico-sociological matter, it does not entail that adaptation should have any genuine place in our normative theorizing or that adaptive preferences are—in Khader's words—worthy of special 'questioning by public institutions'. It does not avoid the skeptical challenge. To see this, imagine what we might discover by looking closely at the phenomenon of adaptive preferences. For instance, we might discover that we should rule out preferences that are inconsistent with the relevant domain's objective metric or that are nonautonomous and so forth. But for each of these possibilities, there are going to be a lot of nonadaptive preferences that are also ruled out. If this is correct, to rule out all instances of preferences that display the relevant feature (whatever it is), there must be an independent argument, independent (that is) of the phenomenon of preference adaptation, to rule out or to look askance at preferences that display



this feature. Without this argument, there could be no grounds for ruling out *nonadaptive* preferences that fall into the relevant category. This independent argument (whatever it is) will be both necessary and sufficient to draw the conclusion to which adaptive preferences have pointed us. But if this is correct, no additional philosophical work can or should be done by the phenomenon of adaptation specifically. Adaptive preferences will then have completely dropped out of any argument or explanation of why some preferences are favored and others are not. But this is the very essence of a philosophical red herring: to continue to focus on adaptive preferences by themselves will either produce an argument insufficient to justify the conclusion—namely, that we should reject the significance of preferences that are nonautonomous—or will rule out far too much—by ruling out not only nonautonomous preferences, but also autonomous adaptive preferences. Thus, even if the phenomenon of preference adaptation forms part of the discovery of important considerations in moral/political/axiological theory construction, it should play no part in justifying such theories and cannot explain why the theories in question ought to take the shape they take.

Second, one might say that I have been holding inquiry into adaptive preferences to too high a standard. ‘It was never the goal’, or so it may be said, ‘to suggest that adaptive preferences should play an explanatory role in theory construction.’ Rather, having considered the paradigmatic cases, we have uncovered an important ‘cause for concern’: when someone’s preferences are adaptive, we should consider them with an eye toward figuring out whether they are problematic, rather than holding that they are problematic and attempting to construct theories around that fact.

However, while the argument of this paper is compatible with such a proposal, to hold that the notion of adaptive preferences was never intended to play a role in theory construction or that no one after all actually accepts the adaptive preference intuition is just wrong. For example, note that Rickard claims that whether we should accept or reject welfarist consequentialism depends on whether it ‘has the resources to legitimately exclude adaptive preferences from counting toward the evaluation of states’ (Rickard 1995: 280). More broadly, Terlazzo provides a helpful rundown of the various ‘strains’ of discussion of the phenomenon of adaptive preferences:

The first of these focuses on whether our preferences are authoritative over our well-being, and aims to explain how and to what extent adaptation to circumstances might call into question the possibility that they are. The second focuses on the way in which patriarchal systems mangle women’s preferences, and aims to explain both the means by which women can come to endorse their own oppression and what exactly is wrong with their doing so. And the third strain focuses on the possibility that living in conditions of severe poverty or oppression can cause persons to have unexpansive preference sets, and aims to explain whether and why such preferences should count in favor of the justice of such circumstances. (Terlazzo 2015: 179)

In all such cases, adaptation is not simply treated as a ‘cause for concern’ but as a reason for thinking such preferences problematic, as ones that should not be favored in the relevant domain. And hence if my argument here is correct, this should be of substantial interest indeed when we come to evaluate the prospects of using adaptive preferences in the manner so specified.

Third, one might suggest that the point I have made here, while (potentially) correct, is nevertheless of merely verbal significance. Surely many, if not most, will agree that some adaptive preferences are not problematic while many nonadaptive preferences are. If so, those I have challenged here might simply respond by saying something like: ‘I simply mean to be referring to *this* set of preferences’, naming the relevant feature, ‘not *all* or *only* specifically adaptive preferences’. And while the argument here might have the effect of demonstrating why they should not use the term adaptive preferences, this is merely a terminological issue.

In response—and building on the previous paragraph—this suggestion seems to me to be uncharitable. To maintain this suggestion one must believe that the adaptive preference intuition is not actually believed by those who explicitly take it as a jumping-off point. As noted by Terlazzo, many take the adaptive preference intuition seriously in crafting accounts of justice, well-being, and morality (at least). Many are interested in explaining why adaptive preferences are (Colburn 2011: 52; Rickard 1995) or are not (Bruckner 2009; Barnes 2009) problematic. If we are to take seriously the aims of these authors, we should not simply hold that whether adaptive preferences are a red herring is a mere terminological issue. Of course, I would be prepared to believe that whether adaptive preferences are a red herring is a terminological issue if there were widespread disagreement about what adaptive preferences refer to. (‘No, what I mean is “autonomous preferences”’; ‘what I mean is preferences that are inconsistent with flourishing’, etc.) But while there is some disagreement along the margins, virtually all who discuss this matter agree with the characterization I offered in [section 1](#). And hence whether to treat the phenomenon of preference adaptation as a reason to hold that particular preferences are nonfavored or as a simple red herring is not merely terminological, but a substantive question when it comes to inquiry into welfare, morality, and political justice.

A final point to conclude this section. In my view, the focus on the phenomenon of adaptation is not just problematic from a theoretic point of view (though that is my main interest), it is problematic when it comes to the fundamental concerns of theorists (or many of them, anyway) who inquire into such preferences. Much of the literature on adaptive preferences seems—for good reason—to be focused specifically on cases in which people suffering from poverty, oppression, sickness, and other terrible circumstances might form preferences we think are morally problematic or unjust and that form a challenge for traditional explanations of why we ought to relieve these circumstances. But if our interest is in responding to the conditions of oppression many people face and many such people maintain preferences that are inconsistent with their basic flourishing, why should we concern ourselves with whether their specific preferences are adaptive? To do so, it seems to me, is a disservice not just to moral theory, but to the oppressed, poor, and sick, as well. Of course, some may respond that some nonadaptive preferences may also be

problematic and should not be taken seriously. Fair enough. But this just establishes that focusing on adaptive preferences—whatever our more general concerns—is a mistake.

#### 4. Conclusion

As noted at the beginning of this paper, literature on adaptive preferences is vast and growing. This is easy to explain. After all, the adaptive preference intuition is very strong. However, as I have argued here, the adaptive preference intuition is misleading. There is nothing that is per se problematic specifically about adaptive preferences—though there may be adaptive preferences that are problematic, just as there are nonadaptive preferences that are problematic in the various ways discussed here. In short, the phenomenon of adaptive preferences does not deserve the attention it has received.

Thus, I think we would do well to expunge the phenomenon of preference adaptation from further consideration of the relevant domains. Adaptive preferences are a red herring; we should focus, instead, on other, more directly explanatory ways of cleaving favored and disfavored preferences (or providing a rationale for ruling out reference to preferences altogether). Jayamma's preference is not problematic because it is adaptive. If it *is* problematic, it is so because it is a preference for unjust or oppressive conditions or a preference that does not capture her autonomy or a preference that does not reflect her true values or a preference that is biased toward the status quo. Whether it is adaptive is entirely beside the point.

DALE DORSEY

UNIVERSITY OF KANSAS

[ddorsey@ku.edu](mailto:ddorsey@ku.edu)

#### References

- Baber, Harriet. (2007) 'Adaptive Preference'. *Social Theory and Practice*, 33, 105–26.
- Barnes, Elizabeth. (2009) 'Disability and Adaptive Preference'. *Philosophical Perspectives*, 23, 1–22.
- Bruckner, Donald. (2009) 'In Defense of Adaptive Preferences'. *Philosophical Studies*, 142, 307–24.
- Colburn, Ben. (2011) 'Autonomy and Adaptive Preferences'. *Utilitas*, 23, 52–71.
- Cudd, Ann. (1994) 'Oppression by Choice'. *Journal of Social Philosophy*, 25, 22–44.
- Dorsey, Dale. (2010) 'Preferences, Welfare, and the Status-Quo Bias'. *Australasian Journal of Philosophy*, 88, 535–54.
- Dorsey, Dale. (2012) *The Basic Minimum: A Welfarist Approach*. Cambridge: Cambridge University Press.
- Elster, Jon. (1982) 'Sour Grapes: Utilitarianism and the Genesis of Wants'. In Amartya Sen and Bernard Williams (eds.), *Utilitarianism and Beyond* (Cambridge: Cambridge University Press, 1982), 219–38.
- Frankfurt, Harry. (1971) 'Freedom of the Will and the Concept of a Person'. *Journal of Philosophy*, 68, 5–20.
- Hawkins, Jennifer. (2008) 'Well-Being, Autonomy, and the Horizon Problem'. *Utilitas*, 20, 143–68.

- Khader, Serene. (2011) *Adaptive Preferences and Women's Empowerment*. Oxford: Oxford University Press.
- Lewis, David. (1999) 'Dispositional Theories of Value'. In Lewis, *Papers in Ethics and Social Philosophy* (Cambridge: Cambridge University Press), 68–94.
- Nebel, J. M. (2015) 'Status Quo Bias, Rationality, and Conservatism about Value'. *Ethics* 125, 449–76.
- Nussbaum, Martha. (1999) 'American Women: Preferences, Feminism, and Democracy'. In Nussbaum, *Sex and Social Justice* (Oxford: Oxford University Press), 130–53.
- Nussbaum, Martha. (2000) *Women and Human Development*. Cambridge: Cambridge University Press.
- Rickard, M. (1995) 'Sour Grapes, Rational Desires and Objective Consequentialism'. *Philosophical Studies*, 80, 279–303.
- Sen, Amartya. (1999) *Development as Freedom*. New York, NY: Anchor Books.
- Sher, George. (1997) *Beyond Neutrality: Perfectionism in Politics*. Cambridge: Cambridge University Press.
- Smith, Matthew Noah. (2013) 'The Importance of What They Care About'. *Philosophical Studies*, 165, 297–314.
- Sumner, L. W. (1996) *Welfare, Happiness, and Ethics*. Oxford: Oxford University Press.
- Taylor, James Stacey. (2009) *Practical Autonomy and Bioethics*. New York, NY: Routledge.
- Terlazzo, Rosa. (2015) 'Adaptive Preferences: Merging Political Accounts and Well-being Accounts'. *Canadian Journal of Philosophy*, 42, 179–96.