

# False Consciousness for Liberals, Part I: Consent, Autonomy, and Adaptive Preferences

---

*David Enoch*

The Hebrew University of Jerusalem

You may have your worries about consent—perhaps it’s not as magical as some people seem to think—and some of the worries about consent are precisely the topic of this article. But the starting point has to be that consent—sometimes, at least in nonproblematic cases—is extremely important. This is true in the private as well as the political sphere. Even if not entirely magical, still consent (or something in its very close vicinity) “turns a rape into love-making, a kidnapping into a Sunday drive, a battery into a football tackle, a theft into a gift, and a trespass into a dinner party” (Hurd 1996: 122). My consent—or perhaps my free, autonomous, informed consent—affects, for instance, what it’s morally permissible for you to do to me. It also affects what the law should say of our interaction, and often, what it does say. Consent is a big deal.<sup>1</sup>

Natalie Stoljar has been extremely helpful both in introducing me to some of the relevant literature and in correspondence, when I tried some of my initial ideas on her. For discussions and comments on earlier versions I am grateful to Rachel Achs, Dani Attas, David Black, Lidal Dror, Jamie Dreier, Nir Eyal, Josh Guetzkow, Ronni Gura Sadovsky, Avi Kenan, Steven Lukes, Bar Luzon, Lije Millgram, Ittay Nissan-Rosen, Massimo Renzo, Mark Schroeder, Rosa Terlazzo, David Weberman, Preston Werner, Daniel Wodak, the editors and readers for the *Philosophical Review*, and audiences in Paris, Montreal, Jerusalem, Harvard, Oxford, Stanford, CEU, Dartmouth, and Berlin. This research is supported by the Israel Science Foundation, grant number 439/15.

1. One may have worries about some of Hurd’s examples, at least in the snappy

But consent is also suspicious, as are its relatives—even its more enthusiastic, positive relatives. For we know that consent is given, choices are made, preferences are formed in specific circumstances, usually from a heavily restricted set of options. And we know that the normative force of a specific utterance of “I hereby consent” strongly depends on these background conditions. For example, even though consent often suffices for transferring property rights, my consenting to the threatening robber’s taking my wallet—indeed, enthusiastically handing it over to him—does not grant him property rights in my wallet or render his use thereof morally permissible. And—much more problematically—the fact that the oppressed sometimes consent to the conditions of their oppression cannot—certainly not always—vindicate the oppression, or render it no longer oppressive. Indeed, as is often noted, sometimes such consent is an indication of just how deeply oppressive the relevant social arrangement is. This is why it is a disturbing (and rare) failure on Mill’s part to vindicate Mormon polygamy, at least when it comes to state coercion: having denounced it morally in very strong terms, Mill (1859: 174) proceeds to rule it illegitimate for the state to forcibly intervene simply because it is “voluntary on the part of the women concerned in it,” as if this settles things, without the need to inquire into the conditions in which the women’s consent is given and their will is shaped.<sup>2</sup>

This complex starting point—that consent matters and that it is suspicious—should be, as far as I can see, *everyone’s* starting point. Its plausibility does not depend on any “ism” in ethics or in political philosophy. Historically, I think it is fair to say that liberalism has been more often associated with taking consent to be a big deal, whereas some other isms (Marxism, for instance) tend to emphasize more the suspicions about consent’s role.<sup>3</sup> But at the end of the day, we should all accommodate both consent’s significance and the suspicions about it. Being a liberal, I am primarily concerned with how liberals can do that, and especially

---

wording that was needed to make this sentence so memorable (and so widely cited). But such worries do not threaten the main point in the text here, I think.

2. The failure I attribute to Mill is not the conclusion he reaches—whether the state should intervene in such cases is a very tricky question, and nothing here seems obvious to me. Mill’s failure is that he fails to notice the complexities and that he treats consent (or the polygamous arrangements being “voluntary on the part of the women”) as if it’s the end of the matter. There is another possible reading of Mill here, which I briefly note below.

3. Feminists are an interesting case, I think, because many of them *have* been sensitive to both aspects of the starting point.

how they can accommodate the suspicions about consent (and avoid Mill's failure). This article is my first step in that direction.

I start—in section 1—with a brief, predictable story about why it is that consent matters. Consent matters, I argue, because autonomy is of value. Section 1 also includes a preliminary discussion of the values of autonomy (I distinguish between nonalienation and sovereignty). In section 2, I discuss the nature of adaptive preferences, and things like consent or choices that may be based on them. It turns out that the very characterization of adaptive preferences needs some work. In particular, we need to distinguish adaptive preferences from other closely related phenomena. With a picture of adaptive preferences at hand, I proceed to discuss whether there is anything irrational about them (in section 3), and whether they suffer from an autonomy deficit (in section 4). The former discussion suggests that adaptive preferences may be preferences formed for reasons of the wrong kind. But the discussion of autonomy here is where much of the action is. Thus, in section 4 I argue that a subset of adaptive preferences—those causally shaped by injustice—do suffer from an autonomy deficit and that this has important moral implications, both in terms of the contribution to the agent's well-being and in terms of the reasons counting against interference. In section 5, I return to consent, and indeed, to politics. The political implications there drawn are very tentative, partly because of complex practical considerations only hinted at here and partly because of the value of sovereignty.

### **1. Why Does Consent Matter?**

I take it as a starting point, then, that consent matters. But even if this is obvious, it is not as obvious *why*. And so it's necessary to spend some time on this question. Having at hand an account of why it is that consent matters (when it does) will be helpful when we try to determine when it does not. So in this section, I put forward an answer to this question. But I want to concede at the outset that the answer will not go very deep—that consent matters may be thought of as very close to moral bedrock, so there's an upper bound on how informative an explanation here can be. My explanation—in terms of the values of autonomy—may at times sound very close to the explanandum. This, I think, is in our context inescapable. Still, it won't be entirely contentless or vacuous, as we will see.

An autonomous life is, other things being equal, a better life precisely for being autonomous. You live a more autonomous life the more control you have over your life, the more major (and perhaps also some

minor) decisions are made by you, in accordance with your deep commitments. The autonomous person is, as is often said, the author of her life story,<sup>4</sup> at least to a considerable extent. That autonomy thus (imprecisely) understood is of value is a very common, and a very natural, thought. If you don't see it, think of the alternatives. A person who stumbles onto a career rather than chooses one (not even in the retroactive, endorsing kind of way); someone who is forced into her relationships rather than chooses them; someone who lives in the only place that was ever an option for him; perhaps also the cult member who follows the leader's extremely specific rules and regulations to the letter, never second-guessing them; there is something important and valuable missing in such people's lives. Their role in their life story is that of a passive participant rather than that of an author. They don't shape their lives themselves; rather, they find themselves living those lives. And while there may be a lot of value in such lives, there is also something deeply troubling about them. They lack on the good of autonomy.

Elsewhere I distinguish between two closely related values that are often associated with the word "autonomy"—that of nonalienation and that of sovereignty (see Enoch 2017a). You're autonomous in the sense of nonalienation vis-à-vis an action or a decision that concerns you to the extent that the relevant matter is determined by your values, or your deep commitments.<sup>5</sup> If you are being subjected to coercive state policies that are based on principles that are inconsistent with the principles you believe in, this is an attack on your autonomy understood as nonalienation. You're autonomous in the sense of sovereignty if you have the last word regarding the relevant matter, if it's a matter of your choice.<sup>6</sup> A paternalistic intervention is always at least *prima facie* an attack on your

4. This metaphor comes from Raz (1986: 369). But see also—in a related context—Christman's (2009: 9) literary critic metaphor.

5. It's important to see that talk of alienation (or nonalienation) here is importantly different from Christman's talk of alienation—for him, alienation is a matter of *feeling* alienated. I go (in this respect, as well as another that will become clearer later on) for a much more objective account. For a related point, see Weberman's (1997: 212–13) rejection of the thought he finds in Engels, that gives pride of place to the *agent's beliefs* about the causal history of his or her preferences. Also, not *any* deep commitments will do for nonalienation, as I argue below. For now, though, this will do.

6. Perhaps there are some things—one's actions, perhaps, or one's intentions—about which one cannot but be sovereign. But autonomy as sovereignty extends much further than that. One can be or fail to be sovereign regarding what is done to one, what options are available to one, and so on.

autonomy understood as sovereignty.<sup>7</sup> Now, often—perhaps even paradigmatically—nonalienation and sovereignty come together. But not always: a policy that is in some sense justifiable to me may be in line with my deep commitments, and so may not be alienating, but if it is politically enforced without my (actual) consent, it may still infringe on my sovereignty. And a paternalistic intervention may—if I am weak-willed, perhaps—make the relation between my life and my deep commitments more rather than less harmonious, but this doesn't mean it doesn't offend against my autonomy in the sense of sovereignty.<sup>8</sup> It seems plausible to suppose that the relation between nonalienation and sovereignty is stronger than mere correlation—perhaps, for instance, at the end of the day one of them is reducible to the other, or perhaps both are best understood in terms of some other related value. For now, though, let me postpone discussion of this possibility for another occasion<sup>9</sup>—whatever the exact relation between nonalienation and sovereignty, the two are distinct, and the distinction between them will be important in what follows.

The literature on autonomy is large and varied, and it's important—in order to preempt misunderstanding—to be explicit about the kinds of autonomy I will *not* be concerned with here.<sup>10</sup>

First, the kind (or kinds) of autonomy that is important here—that is a part of or is closely related to what makes a life go better—has nothing to do with an ideal of individuals as small autarkies. Relations

7. I discuss—in a *very* preliminary way—the relation between paternalism and autonomy in my essay “What’s Wrong with Paternalism” (Enoch 2016).

8. Cases in the opposite direction—where there is an offense against a person's autonomy-as-nonalienation, but not against their sovereignty—are possible, but things are more complicated here, so it's harder to think of a clean, uncontroversial example. Here's a close one: suppose a person is deeply committed to norms of sexual modesty, but finds herself weak-willed on a specific occasion, wanting (and intending) to have sex with the man she's having a drink with. Perhaps—but again, things are murky here—if at her initiative he has sex with her this offends against her autonomy-as-nonalienation, but it most certainly doesn't offend against her sovereignty. (Perhaps if he declines only for the reason that having sex is not in line with her deep commitments, then *this* would offend against her sovereignty.)

9. As I did when first introducing the distinction.

10. Thus, it may be thought that “autonomy” is ambiguous. Killmister (2018: 5) wants to avoid this move, insisting that a good (multidimensional) account of autonomy will cover almost all of the relevant uses. I reject such a condition if it's understood as an adequacy constraint on an account of autonomy, but I remain open to the possibility that substantive argumentation later on will show more unity than we are entitled to at the start.

with others are—quite obviously—crucial for (at least most ways of) living a good life, and also, at least for most of us, for living an autonomous life.<sup>11</sup> Nothing in the value of autonomy as sketched above idealizes utter independence, and the idea that this is what liberals are after when they value autonomy seems to me to be a characteristic more of caricatures of liberalism than of the real thing.

Second, as far as I can tell, there's nothing Kantian going on here at all. For Kant, being autonomous is—roughly—about being subject to only laws that one legislates oneself, and while I can say these words, it is not at all clear to me what they mean.<sup>12</sup> But however understood, a commitment to the value of autonomy as sketched above need have nothing to do with such an ambitious idea. As far as I can see, for instance, there's no tension at all between the thought that one is subject to moral laws that one does not in any way legislate and the thought that there's value in being an author of one's life story.<sup>13</sup>

Third, some discussions of autonomy tie it very closely—perhaps with Kantian inspirations—to metaphysical discussions of free will and to the related ethical ones of moral responsibility.<sup>14</sup> The perhaps natural thought is that acting autonomously is a necessary condition for acting freely, or for being responsible, or is in some other way closely related to freedom and responsibility. This discussion too is distinct from mine. At least initially, we can't rule out the possibility that a life manifests the value of autonomy as described above, and yet that freedom of will and moral responsibility are an illusion. Similarly, it's not impossible to think of someone choosing—in a way that renders them fully responsible—to live a much less autonomous life (this is one way of thinking of the cult member, for instance).

Of course, both with regard to Kantian autonomy and with regard to the kind of autonomy purportedly relevant to free will and responsi-

11. Whether this takes us all the way to “relational autonomy” is a different question, one that I can remain undecided on here. See Oshana (1998); and Mackenzie and Stoljar (2000).

12. For the claim that this idea is both empirically and conceptually suspicious, see Dworkin (1988: 36).

13. Fortunately for me, as I am a moral realist, and I find no place in my metaethics for anything resembling Kantian self-legislation. See my *Taking Morality Seriously* (Enoch 2011b); and for critiques of what are arguably more Kantian-friendly views, see my “Agency, Shmagency” (Enoch 2006) and “Can There Be a Global, Interesting, Coherent Constructivism about Practical Reason?” (Enoch 2009).

14. See, for instance, Buss and Westlund (2018) and the references therein.

bility, we can't rule out from the start the possibility that a deeper understanding will reveal close connections between them and the values of autonomy (in terms of sovereignty and in terms of nonalienation).<sup>15</sup> But first, such connections will have to be established; we cannot just assume them, or pretend that we've supported them by tendentious, perhaps equivocating uses of the word "autonomy." And second, even if deeply related, they are distinct, and for what follows, this will suffice.

Let's get back to consent. Consent matters because autonomy is of value. We have the normative power to change—by consenting—the normative landscape, to render otherwise impermissible actions permissible (or less impermissible), because having such a power is necessary for living an autonomous life. A creature whose consent doesn't make a huge difference to the moral status of a sexual interaction with it is not the author of the sexual parts of its life, and because these are such central parts of life (for creatures like us, at least), its ability to lead an autonomous life is severely restricted. A person whose consent doesn't affect what may permissibly be done with some physical aspects of her surroundings does not in fact have property rights, and assuming that having some control over some property is important for leading an autonomous life (at least in an environment rather similar to ours), her ability to lead an autonomous life is seriously compromised. And so on.<sup>16</sup> The values of autonomy, then, are closely related to the power of consent. The value of autonomy as nonalienation—of living one's life according to one's own deep commitments—partly explains why we possess the power of consent. When it comes to autonomy as sovereignty, the relation may be too close to be explanatory—possessing the power to change the normative landscape with one's consent and being sovereign in the sense of having the final word on some topic are close to being identical.<sup>17</sup> But, first, they

15. See Buss (2005: 5) for the claim—without support—that the notion of autonomy she's working with is related to the political one.

16. All of this, notice, is about the *normative* force of consent. One may also ask questions about the nonnormative power of consent. Arguably, a creature whose consent does make a difference as to what may *permissibly* be done to her, but doesn't make any difference as to what *is likely* to be done to her, also suffers from a huge autonomy loss. But it's a different loss from the one in the text. Perhaps one of the ways in which autonomy is relevant to the power of consent is if, in circumstances that are not too bad, there are close relations between how one is likely to be treated and how one may be permissibly treated. But autonomy seems relevant—in the way described in the text—even in the absence of such correlation. The emphasis on normative interests of the kind discussed here is central to Owens (2012).

17. I thank an anonymous referee for this observation.

are not *quite* identical. Consenting to something may render an otherwise impermissible action permissible without yet settling whether or not it will (or ought to) be done. And second, insisting on the role of sovereignty here remains important, if only as a reminder of the negative claim that nonalienation is not the full story underlying the normative power of consent. Still, it will be the part of the story I focus on here.

This explanation of why you possess the power to change the normative landscape with your consent refers to the value of autonomy, that is, the value of shaping your own life. This explanans is, as I warned, rather close to the explanandum. But it's not as if no progress at all has been made. For one thing, it seems like a plausible conjecture that *all* normative powers are explained in terms of values (Watson 2009: 162): the normative power you possess to give me a reason for doing something by requesting that I do so is plausibly explained in terms of the value of our relationship (and the fact that possessing such a power is a constitutive part of such a relationship; see Enoch 2011a); the fact that we have the normative power to obligate ourselves by promising is plausibly explained by noting how a world in which we have such a power is better than worlds in which we do not, and so on. If so, it's an advantage of the story just told that it fits this wider picture. And second, while the relation between consent and the value of autonomy is natural, it is not entirely trivial—as will emerge later on, when it is put to use in drawing substantive implications.<sup>18</sup>

If consent matters because autonomy is of value, it is very tempting to think that consent only matters where the value(s) of autonomy is (or are) at stake. This is indeed a very plausible thought, but as it stands it's nonetheless false. By and large, consent matters because autonomy is of value, but perhaps the power of consent—based as it is on the value of autonomy—acquires a life of its own, and retains some of its significance even when the value of autonomy is not relevant. This is possible, for instance, if—in a given social context—the power of consent becomes the canonical manifestation of the value of autonomy, perhaps symbolizing autonomy even when autonomy is not present,<sup>19</sup> or if the justificatory

18. And as is perhaps also evidenced by the fact that this relation has recently been called into question. See Healey n.d. I cannot here address Healey's arguments, but let me note that even if he is right, much of what I am about to do in this article can be done in terms more friendly to his view about the grounds of the power to consent.

19. Raz (1986: 378) argues that the badness of coercion in cases where it does not compromise autonomy (as he understands it) can be explained in a similar way.



structure here is in some other way indirect.<sup>20</sup> The details may be tricky—sometimes it may be hard to distinguish this kind of move from some irrational means-fetishism. But these are possibilities that it will be important to keep in mind.

The natural story about consent mattering because autonomy is of value is consistent, of course, with consent sometimes mattering also for other reasons, instrumental ones chief among them. From the point of view of the moral engineer, as it were, there are often very good instrumental reasons to grant people the power to change the normative landscape with their consent<sup>21</sup> (think of all the benefits of trade, for instance). It's just that such instrumental reasons are contingent, and they don't exhaust the significance of consent. The central, to a large extent necessary, intrinsic, reason consent matters, it seems to me, is the autonomy one.

## 2. Adaptive Preferences

As noted at the outset, accommodating and explaining the ways in which consent is suspicious is a central aim of this article. Often, when the literature addresses such suspicions, it does so in the context of discussing adaptive preferences. It will prove useful, then, to spend some time on those, starting with examples.

The fox really wants to have the sweet grapes, but having found out that they are out of his reach, he loses that preference, perhaps convincing himself that they are not that sweet after all.<sup>22</sup>

A teenager shows some talent in music and in philosophy. What she would want to do most is to compose great music. With time, though, it becomes clear that she is nowhere nearly talented enough to do that.

20. Even if you think, for instance, that freedom of expression is ultimately grounded in the effect on finding out truths, you may think that it protects some instances of speech that are in fact counterproductive to that end.

21. At least, assuming the kind of correlation mentioned above, in footnote 16.

22. This example is based on Elster's (1983) use of La Fontaine (though this specific fable is more often attributed to Aesop). The example is now canonical—you can't really discuss adaptive preferences without it. But, as will emerge, it's not really that good of an example after all. The term "adaptive preferences"—coined by Elster with this canonical example—has developed in ways that render the example less than ideal. It's not even clear that Elster's original understanding of the term applies to the oppression cases—central ones in the contemporary literature on adaptive preference. See Elster (1983: 128–29).

For the point that the term "adaptive preference" is now used much more broadly than in Elster's text, and that oppression cases have come to be central to the use of this term, see Terlazzo 2016: 206, and the many references therein.

She can, however, be a decent philosopher. So she chooses that path, and—perhaps with time—comes to prefer it over others, the music one included.

A young woman grows up in a highly conservative society—perhaps in South Asia<sup>23</sup>—where it is common for women to malnourish or even starve themselves in order to better feed (well beyond need) their husbands and male children. And she chooses to malnourish herself in this way. Gender norms play a role here, of course, but at the end of the day, the woman prefers to have her husband eat more than enough rather than have enough for herself.

These examples (a couple of others will emerge as we proceed) are rather typical of what has come to be called “adaptive preferences.” But they differ from each other in important ways, and it’s crucial to proceed here more slowly, for even the characterization of adaptive preferences is far from unproblematic.

Before giving more details, let me just say that while it is now standard to speak of adaptive *preferences*, preferences are not that special in our contexts. I will be just as concerned with adaptive *choices*, adaptive *decisions*, adaptive *consent*, and the like.<sup>24</sup> Depending on your theory of motivation and philosophy of action, you may think of all these other cases as somehow parasitic on the preference case—perhaps the best way to think, for instance, about an adaptive choice is as a choice motivated by an adaptive preference. Whether such a claim is plausible will depend on wider issues than I can discuss here, among them on how best to understand “preference.” Because I want to remain neutral on such matters, I’m not going to be making such assumptions. However the relations between preferences and choices and decisions and consent (and perhaps other things) are best understood, the distinction between those that are adaptive and those that aren’t holds, in roughly the same way. And below, when I speak of adaptive preferences, this should be understood as shorthand for the wider phenomenon of which these are a subset. Let me just draw atten-

23. The example is based on Stoljar’s (2014: 237) discussion, itself based on reports by Khader (2011).

24. For a similar point, see Walker 1995: 470n1. A *really* interesting kind of case is that of adaptive *hedonic states*. I’m not sure what to think of those, perhaps because it’s anyway not clear how to think of autonomy in the context of hedonic states. (Notice that adaptive hedonic states in the sense I’m using the term here are distinct from the phenomenon sometimes called “hedonic adaptation.” For discussion of hedonic adaptation in a closely related context, see Mitchell 2018. And for an interesting suggestion about the relation between hedonic states and the adaptation of preferences, see Millgram 2000.)

tion to a context in which talk of preference may be especially misleading. When the woman in the third example supports her community's practices, this may be a matter of preference, but it may be something quite different, perhaps her values, or a commitment to a set of principles. And while there may be a sense of "preference"—perhaps the choice-theoretical sense—in which it's OK to speak of preferences even in cases of action motivated by value judgments or a commitment to principles, this is a sense far removed from the natural-language one.<sup>25</sup> So it's important to remember that the discussion that follows applies also to adaptive commitments to values and principles.

How, then, should we understand adaptive preferences? At a first approximation, "adaptive preference is a preference that is regimented in response to an agent's set of feasible options" (Bruckner 2009: 308). The food preferences of the fox change not because of a change in the relevant information, or the nature of the grapes, or indeed the kind of change in one's palate that sometimes occurs, though typically over time (I used to *just love* quinoa salad, but it now leaves me cold). Rather, it's a response to the fox's realization that consuming the grapes is not a feasible option for him. Similarly, nothing made the prospects of writing great music less appealing, and it's not as if the teenager in our example had the kind of change of heart that is in some sense more standard. Rather, she realized that that option is not in the cards. And if you think of the case of the woman who is malnourishing herself analogously, then you probably think of some other options—having a more egalitarian division of resources, living in a society where there are no gender-related norms that call for such discrimination, and so on—as options that are in some sense better for her, and for which she lacks a preference just because of the kind of conditioning she's been the victim of, or the way the oppressive environment that is the only environment she knows has shaped her preferences.<sup>26</sup> (It is not *at all* obvious that this is how this case is most naturally thought of. I return to this below.)

25. I highlight one of the perils—in political philosophy—of thinking of commitments to principles as mere preferences in my "Against Public Reason" (Enoch 2015: 131).

26. In this example too I could have put things in terms of feasibility: the option that would be best for her would be to continue to be an insider (with all the benefits this includes), without the need to malnourish herself. This option is rendered unavailable by the patriarchal order in her community. But there's something a little forced in this way of presenting things, and anyway, I don't see what's gained by doing so.

But, for the category of adaptive preferences to be theoretically productive, we should proceed beyond this first approximation. Several points are important here.<sup>27</sup>

Some preferences are based on false beliefs, and, when they are, this seems to be relevant to their normative relevance. It's not obvious *in what way* exactly being based on a false belief affects the normative relevance of the preference—perhaps it renders the preference non-reason-giving, or perhaps less so, or perhaps the relation to a false belief is in some other way relevant. This—the possibility that a preference is based on a false belief—is, of course, a very wide phenomenon, and it has nothing necessarily to do with adaptive preferences.<sup>28</sup> Many classical false-consciousness cases are of this kind (so an article about adaptive preferences cannot exhaust the topic of false consciousness): a low-income Walmart employee who believes he lives in the Land of Unlimited Opportunity even though no one in his family has ever gone to college and in all likelihood no one ever will, and who bases his political choices on these beliefs, is clearly motivated by false beliefs (or at least ignorance of some true ones). But think about our fox again. One natural way to complete the account of his psychology is to add a false belief—namely, the belief that the grapes are sour. If so, the fox's preference revision becomes a standard response to a change in beliefs, and then, it may be thought, it's no longer in any interesting sense an adaptive preference. It's just a preference based on a false belief. I think there's something to this thought, but it's too quick. True, the shift in preferences is based on a change in beliefs, but *that* change is not based on, say, new evidence, but rather on something more like wishful thinking: the unattainability of the grapes is no evidence that they are sour; at most it's reason to *hope* that they are. So while the fox's new preferences are indeed based on false beliefs, they are not merely nonspecial particular instances of the set of preferences based on a false belief. Rather, they are preferences based on false beliefs themselves acquired as a (nonevidential, perhaps irrational) response to a realization about one's feasible options. There is still a clear

27. "Adaptive preference" is a technical term, so we don't have pretheoretical intuitions about it. So the method I will be employing is—as hinted in the text—finding out how adaptive preferences are to be understood if it's going to be an interesting fact about a preference that it is adaptive. (Another way to go here would be to stipulatively define the term such that, for example, adaptive preferences are guaranteed to be irrational; cf. Mitchell 2018. But it's hard to see how such stipulations can promote understanding.)

28. This is a major point for Bruckner (2009), though he puts it more strongly, and he doesn't note the point that follows in the text, about the relevant beliefs being adaptive.

sense in which the fox's new preference is adaptive, then. (It's not clear to me whether something similar may be said of the Walmart employee.)

Now think of the music-loving teenager again, and assume that had she been asked—before the painful realization that she's not going to be a composer—to rank career options, she would have ranked composer first, philosopher second, and lawyer third. Now the evidence indicates that composer is out. She may still prefer a career in philosophy over one in practicing law. There is nothing in itself problematic—or as far as I can see in any useful sense adaptive—about *this* preference. After all, the comparative preference between philosophy and law remained stable, even when the evidence about the infeasibility of a career in music became apparent. In what sense, then, is her preference for philosophy (or her choice to pursue a career in philosophy) adaptive?<sup>29</sup> Is it possible that the temptation to classify her preference as adaptive is entirely due to our sloppy neglect to mention what she prefers philosophy *over*?<sup>30</sup> I think that if classifying the preference for pursuing philosophy as adaptive is to be at all theoretically productive, it must be understood as a preference for a career in philosophy *over one in music*. If this is the relevant preference, then we see how it is formed as a response to a change in the perceived set of feasible options, and we also see how it's at least a little suspicious. The teenager has to convince herself somehow (consciously or unconsciously<sup>31</sup>) that *even had she been able to become a great composer, she still (now) prefers pursuing philosophy* for her preference for philosophy to be interestingly adaptive. Compare: a fox who steps out of the vineyard with

29. Notice that this kind of example is a paradigmatic example in the literature on adaptive preferences. See, for instance, Nussbaum 2001: 78–79, and Bruckner 2009: 313. And yet I don't know of a text that addresses the problem in the text here. Most of Bruckner's (2009) discussion, for instance, is unclear precisely on this kind of problem.

30. Elster (1983: 131) is more careful on this than those who have followed him. See Railton n.d. for the claim that people can—and as empirical research shows, do—have higher degrees of subjective well-being with, say, permanent colostomy sacs than they had anticipated, and indeed, higher than patients with temporary colostomy sacs (so that some habituation occurs), but retain the preference for a life without one (had that been possible). Such cases are not, then, cases of adaptive preferences in the sense relevant here. And of course, there need be nothing irrational about such a preference and subjective well-being structure.

31. Elster (1983: 117) emphasizes this distinction, reserving the term “adaptive preference” to only unconscious changes of this kind, while calling conscious changes “character planning.” He thinks the distinction is important, because (among other reasons) character planning is more autonomy-friendly than adaptive preference. This seems to me to be an oversimplification, but there is a truth in the vicinity here—see the discussion of making a preference one's own, below.

the quiet understanding that he can't get the grapes, and because his second-best option is to eat something less appealing rather than to keep staring at the grapes and starve, is not manifesting adaptive preferences (also, it's a fox that can continue to acknowledge the grapes' sweetness). It becomes a case of adaptive preferences only when the fox convinces himself that he *wouldn't want the grapes even had they been attainable* (because they're sour).

Things are tricky here, because feasibility considerations may straightforwardly affect the desirability of an option, and then, taking such effects into account will not render the preference adaptive in the relevant sense. If the fox finds out not that the grapes are literally unattainable, but that attaining them will have a great cost (say, in physical fatigue, perhaps risk, etc.), then he may perfectly rationally abandon the attempt to get them, even without (perhaps irrationally) forming the belief that they are sour. In such a case, all that's going on is that the fox gets new information about the costs (in terms of other standing preferences of his) of the grapes. He still prefers having the grapes over not having them, if all else is equal; he just understands that it's not the case that all else is equal. Similarly, for the music-loving teen, if she finds out that the costs of being a composer (say, in terms of personal relationships) or of the attempt to become one (say, in terms of risk) are much higher than she had thought. For the case to be interestingly classified as an adaptive preference change, the change of preference (say, of philosophy over music) must have occurred independently of such costs of the relevant options or of trying to get them. And this means that clean adaptive preference cases are going to be harder to find than we may have thought, and that it's going to be potentially controversial, for each purported adaptive preference case, whether it is one.

Return now to the example (from Mill) of the woman who voluntarily takes part in a polygamous marriage in the Mormon community. We can now see that things are not as clear-cut here as we may have thought. First, it's possible that her preference is at least partly the result of a host of false beliefs that are not the result of adaptive mechanisms. Second, we need to answer the comparative question—what is her preference (for becoming, say, the third wife of a certain man) a preference *over*? If it's merely preferring that option over that of remaining unmarried in the Mormon community, then the preference may not only be (depending on complicated social and personal factors) perfectly rational, but also not adaptive in any interesting sense. If it's the preference of that option over the (unattainable, but assuming it was attainable) option of living in

a more egalitarian social order, then understanding the preference as adaptive seems more plausible. And even then, if she prefers, say, to stay a part of the Mormon community because of the extra costs of realizing some fairly abstract “exit right,” then this preference too may be perfectly rational and in no interesting way adaptive—she may just be taking into account the relevant costs of leaving (costs unjustly imposed, and without which perhaps she would have preferred to leave). Regarding any specific case, then, the question whether the relevant preference is adaptive is not at all straightforward.

These observations are extremely significant when it comes to the political implications of the woman’s preferences and choices. There will be more of this below, but for now: if the woman’s preferences are (merely) the result of false beliefs, perhaps we should treat them in the general way in which we should treat preferences based on false beliefs. It’s not obvious to me what this general way is, and furthermore, perhaps it sometimes matters how the relevant false beliefs were formed (they themselves may be the result of oppression, as is perhaps the case in the example of the Walmart employee). If the woman’s choice is simply the best one among the options available to her,<sup>32</sup> then robbing her of even this (bad, but best available) option, without addressing the underlying factors that make it the case that this is her best available option, will amount to adding insult to injury.<sup>33</sup> And of course, any kind of political intervention will also have

32. This is a crucial distinction. When an agent opts for the best among bad options, we can easily think of his preferences (as between the available options) as entirely his own. When preferences are adapted, this is not the case. Of course, none of this means that preferences may be adaptive without any material influence on them (and on the account I’m going to end up giving, a lot depends on the causal history of the relevant preferences, so obviously material conditions are very much in). It’s just that there is a difference between the material conditions shaping just the options available to the agent, and them shaping the agent’s preferences (or perhaps the agent him- or herself).

33. See Stoljar’s (2014: 231) discussion of “bargaining with patriarchy,” and the references therein (the phrase comes from Uma Narayan).

Mill’s (1859) discussion may be read along these lines: “This relation is as much voluntary on the part of the women concerned in it, and who may be deemed the sufferers by it, as is the case with any other form of the marriage institution. . . . It has its explanation in the common ideas and customs of the world, which teaching women to think marriage the one thing needful, make it intelligible that many a woman should prefer being one of several wives, to not being a wife at all” (174). Here Mill may be making two distinct points, each much better than the point I attributed to him at the outset (that the Mormon women consent, and that’s the end of the matter). First, he may be merely emphasizing the comparative point, namely, that it’s not clear that the polygamous Mormon arrangement is significantly more problematic than the monogamous arrangements his readers

to cope with almost insurmountable epistemic obstacles here: it is going to be very hard to know which description fits the case, and what may need to be involved in finding out may make the inquiry morally unacceptable. I return to these issues below.

In what follows, then, when I speak of adaptive preferences what I have in mind is the kind of preference highlighted by the examples we started with, except after having been cleaned up of the distorting factors noted above. That is, adaptive preferences are preferences<sup>34</sup> that have been shaped or revised in light of information about feasibility, among the same options (so not merely by preferring the second-best option over the third, once it's clear the best option is infeasible),<sup>35</sup> and when the feasibility considerations do not directly affect desirability (as when they raise the price of some option).

### 3. Is There Anything Irrational about Adaptive Preferences?

Thus understood, then, is there anything irrational about adaptive preferences? This question is, I think, of independent interest. But it is also relevant to questions about the autonomy deficit such preferences arguably suffer from (and that I discuss in the next section). This is so because it is often thought that sufficiently serious, sufficiently systematic, sufficiently deep rationality failures undercut autonomy. If adaptive preferences are necessarily (pro-tanto) irrational, this may appear to be a reason to think they also suffer from an autonomy deficit. In this section,

---

are likely to take for granted. So Mill may be here not so much lenient on Mormon polygamy, as revolutionarily harsh on traditional, monogamous Christian marriage (a reading that is perhaps more in line with others of his texts). Second, Mill seems to note that however unjust the background structure, still, *given* it, entering the (polygamous) marriage may be the woman's best option.

This is still not good enough, for a reason I get to in section 5. But it's better. As always, exegesis is hard. It is perhaps possible—as Amy Schuster tried to get me to see—to read Mill as even better, perhaps as addressing the Mormon women with some of his criticism, trying to engage them in reasoning. I don't know if this is so, and I'm, of course, not committed to any exegetical claim here.

34. I don't see what is to be gained from refusing to classify them as "real" preferences. They are preferences, alright, not merely an agent's beliefs about what he or she prefers (they certainly function like preferences and not like beliefs about preferences), except that they are a special class of preferences, about whose rational and autonomous status we are about to inquire.

35. Bargaining with patriarchy cases can be seen as a particular instance of such cases as in the text—for in those cases, the patriarchy severely restricts the set of available options, and the relevant preference is just among the remaining options.



though, I argue that this is not so—I diagnose the way in which adaptive preferences are (pro-tanto) irrational, and then argue that this irrationality does not entail an autonomy deficit, thus clearing the way for the discussion of autonomy in section 4.

The first (rather obvious) thing to note here is that often adapting is, all things considered, a very good idea indeed. The person who keeps aiming at unattainable ends seems clearly irrational. And sometimes, it seems safe to speculate, it will be very useful—and sometimes extremely important—to pursue second-best options as if they were the best.<sup>36</sup> Perhaps, for instance, the music lover can have a much better life if she comes to prefer a career in philosophy over even one in music, had becoming a great composer miraculously become a viable option. And when it comes to the psychological struggles and well-being of those living under oppression, I don't think many useful generalizations can be made, but I'm confident that sometimes convincing oneself that one really does want to have the bad but best available option is an important survival mechanism. In this way, then, adaptive preferences may be perfectly rational.

Some are suspicious of adaptive preferences because they are not rationally formed, and—perhaps relatedly—because their forming mechanisms are often subconscious. This is unconvincing, though—neither being formed rationally nor being formed (or revised) under conscious regulation are requirements for a preference to have its rational credentials. After all, many of our preferences (and choices, and decisions)—perhaps all—are at least partly shaped by causal forces of which we aren't aware. And so long as preferences—or changes in preferences—are not *irrational*, their being at least partly *a-rational* is neither here nor there. Adaptive preferences are in no way special in this regard, and so are not especially suspicious either.

Still, there is a way in which adaptive preferences are—as a matter of necessity, I think—pro-tanto irrational. The paradigmatically rational way for preferences to change is in response to changes in the desirability

36. Bruckner (2009: 314) cites some relevant empirical evidence. Note that the sentence in the text holds even if there's always *something* problematic—perhaps not fully self-aware in a problematic way—in pursuing the second-best option as if it were the first. The point here is about overall usefulness. For an example of a relevant tradeoff, see Killmister's (2018: 99) plausible suggestion that in many oppression cases, adapting will achieve an increase of local autonomy at the expense of a decrease in global autonomy (I say more about these terms later on).

of the relevant object<sup>37</sup> (and sometimes, in response to changes in beliefs or evidence about such desirability, or about the features that make it desirable). Had the music lover's career preference changed because of new information, say, about how much easier it is to make a living as a philosopher than as a musician, this would have been a rational change in preferences, a response to changes in (perceived) desirability. In such a case, the relevant reasons (for preferring, or desiring, or choosing, or . . .) are, in Parfit's (2001) sense, *object-given*—they are about the thing desired. The things that explain the shift in preferences in my story, though, are different. They are not about the desirability of the object desired, but about the desirability of the desiring state.<sup>38</sup> They are about the consequences of being in the desiring state—say, that continuing to prefer a life in music over a life in philosophy will lead to a less happy life— independently of how desirable the objects of the relevant desires are. They are, in Parfit's terms, *state-given* reasons.

Some people think of state-given reasons as reasons *of the wrong kind* for that state. And you can see why: epistemic reasons for belief are, at least in the paradigmatic cases, evidence for the truth of the content of the belief, not anything about the utility of so believing. Good, right-kind-of reasons to want something have to do, at least in the paradigmatic cases, with the desirability of the thing desired, not with how better or worse off the desiring itself will make you. Reasons (of the right kind) to fear something have to do—at least in the paradigmatic cases—with that thing's dangerousness, not with the benefits or costs of being in a fearing state. If someone offers you a lot of money in order to believe that the number of stars in the universe is even, or in order to want saucers of mud, or in order to fear kittens, they may be offering you reasons for those things, but not of the right kind—the money is about the state, not about

37. There are other options that are not, I think, *as* central, and even if they are, this doesn't affect my argument in the text: Perhaps, for instance, with regard to some preferences it's rationally permissible for them to change for no reason at all. And perhaps sometimes preferences rationally change as a result of a shift in attention as between the many different things that may each be individually desirable. (I thank Lije Millgram for this point.) But such cases too are different from the one of adaptive preferences.

38. If preferences are what *determines* desirability, then what I say in the text cannot be true as stated. But first, what I say in the text is very plausible, and so a reason (among perhaps other powerful ones) not to think that preferences determine desirability. Second, even if preferences do determine desirability, not all of them do (perhaps only the basic, nonderivative ones do), and so the point in the text still holds about those that don't. And third, I believe that the story in the text here may be applied—once suitably revised—even to desirability-determining preferences.

the object. It does nothing to show that that belief is true, or that saucers of mud are desirable, or that those little things are dangerous.

If so—if state-given reasons are the wrong kind of reasons—then adaptive preferences are pro-tanto irrational because they are formed or revised under the influence of reasons of the wrong kind (or because they fail to respond ideally to all and only the reasons of the right kind).<sup>39</sup> Preferring a life in philosophy over one in music simply because one can't be a great composer (and because continuing to want this unattainable thing has serious costs)—without any new evidence regarding the comparative desirability of both careers—will be analogous to believing that the number of stars in the universe is even, for the money.<sup>40</sup> Or indeed, it will be akin to believing that the grapes are sour simply because this makes not getting them less painful (but without any evidence indicating their sourness).<sup>41</sup>

I think that often this is what is going on—that often adaptive preferences *are* rationally suspicious, and perhaps also pro-tanto irrational, and that the suspicion is often best diagnosed in terms of the wrong kind of reasons. But things get complicated because the identification of state-given reasons with the wrong kind of reasons is controversial (Schroeder 2012), and because weighing in on this controversy is not something I can do here. Let me just acknowledge this complexity, then, and insist that

39. Whether there's sense to be made of an all-things-considered judgment, one that commensurates wrong-kind and right-kind reasons and delivers one final rationality judgment, is a matter I need not commit myself on here (though I should say I suspect not).

Earlier I said that when adapting is an important survival mechanism, it may be perfectly rational. Here in the text, I note a way in which it may nevertheless be irrational. This is not a contradiction: adapting may be pro-tanto irrational in one way, but not in another, and if there is a relevant notion of all-things-considered rationality, it may be all-things-considered rational as well. In fact, this is a common phenomenon in the vicinity of reasons of the wrong kind. If I offer you a lot of money to believe that the number of stars in the universe is even, there's a sense in which it's perfectly rational for you to accept the offer (if you can). But there's also a sense in which so believing would be irrational—it would be believing for reasons of the wrong kind, against the evidence, etc.

40. A point missed by Richardson (2001: 292), who discusses adaptive preferences in a related context, and treats the advantages of adapting as perfectly ordinary reasons to have the relevant adaptive attitudes.

41. In this way, then, the case of a preference based on a false belief of the kind relevant here is even closer to other cases of adaptive preferences than that highlighted in the text above, in section 2. The fox's belief is suspicious in precisely the same way adaptive preferences are suspicious—they are formed based on reasons of the wrong kind (for belief, and for preference, respectively). Elster (1983: 124) notices the relation to wishful thinking.

even if some state-given reasons are of the right kind for the relevant state, still in many paradigmatic cases this is not so. And this suffices, I think, in order to diagnose the suspiciousness of many adaptive preferences. (If those accepting some state-given reasons as reasons of the right kind win the day, this means that perhaps in some cases of adaptive preference, the initial suspicion can at the end of the day be dispelled.)

Of course, even if adaptive preferences are pro-tanto irrational, it's not yet clear what follows from this.<sup>42</sup> One thing that clearly doesn't follow is anything about the blameworthiness of the person having those preferences. Even if there is always something suspicious about responding to reasons of the wrong kind, it doesn't follow that this is never the all-things-considered rational thing to do, and even when it isn't, it doesn't follow that the person responding unjustifiably is blameworthy—whether this is so will depend on further details. (Perhaps, for instance, given the manageable costs of continuing to admit that being a great composer would have been much better, the music lover is rationally criticizable for having convinced herself that she now prefers a career in philosophy over one in music; but perhaps, given the *unmanageable* costs of thinking of one's marriage and all that comes with it as an oppressive arrangement, the Mormon woman in Mill's example is not criticizable for having the needed adaptive preferences.)

Another thing that doesn't follow is that it would be best not to have adaptive preferences. Indeed, state-given reasons, when they are weighty enough, falsify precisely this claim. The psychological engineer, as it were, may have very weighty reasons—having to do with the costs of having all sorts of preferences, not with the desirability of the relevant objects—to implant in us a mechanism that disposes us to have adaptive preferences on some occasions.<sup>43</sup>

Perhaps most important for our context is that nothing about autonomy follows, it seems, from the way in which adaptive preferences (arguably) respond to the wrong kind of reasons. Preferences shaped by state-given reasons, indeed by the wrong kind of reasons, may be no less autonomous—

42. Walker (1995: 465) thinks the rational standing of adaptive preferences matters greatly for their political significance. I'm not so sure. Note also that in his survey of ways in which adaptive preferences may be thought to be irrational, the wrong-kind-of-reasons way is not mentioned.

43. At least, this is so *given* our imperfections. Perhaps the engineer tasked with designing an ideal agent would have no room for such adaptation—perhaps the ideally rational agent will not suffer from the costs of acknowledging that his best remaining option was his second-best, and so on. I thank Preston Werner for this point.

no less the agent's own, no less an attempt to self-author one's life story—than other preferences. Fully supporting this claim requires a fuller discussion of autonomy, of course, a discussion we are about to turn to. For now, though, let me just note that this claim—that the way in which adaptive preferences are arguably pro-tanto irrational does not amount to an autonomy deficit—follows not just from the account of autonomy I develop below, but also from pretty much all other accounts<sup>44</sup> that I survey in the next section.

#### **4. Are Adaptive Preferences Nonautonomous?**

A major concern in the literature on adaptive preferences is whether they suffer from an “autonomy deficit” (Stoljar 2014). Focus now on the woman who prefers that the men in her family be better fed, at the expense of suffering malnutrition and perhaps even starvation. It does seem like this preference of hers—having been formed in the context in which it was formed, under the relevant pressures, and so on—is not fully hers. One is tempted to use metaphors again—the relevant preference seems to be alien to her, a preference implanted in her by the patriarchal order, rather than something coming more authentically from her. By acting on such preferences, the woman does not author her own life story; rather, she is defeated by external forces. The question is to what extent such natural thoughts can be made more precise, and whether they can be applied more widely.

In order to face this challenge, we need an account of why it matters whether such preferences have an autonomy deficit. As we will see, this may matter in any number of ways, and it's quite possible—and will emerge as important—that adaptive preferences have an autonomy deficit vis-à-vis some of these, but not others. This, then, is one set of relevant distinctions. Another addresses not the implications of declaring adaptive preference less than fully autonomous, but that which makes them so. Let me start with the latter.

##### *4.1. What Makes (Some) Adaptive Preferences Less Than Fully Autonomous?*<sup>45</sup>

The fact that relevant preferences are adaptive doesn't seem to compromise the sovereignty manifested in decisions and actions based on them.

44. With the possible exception of a content-based account. See below.

45. Because I am here asking about a preference (rather than a life) whether it is autonomous, my discussion will be of what has come to be called *local* rather than *global*

Regardless of the nature of the preference, it's still the person whose preference it is who gets the final word on the matter. Indeed, sometimes taking a preference less seriously because it is adaptive will amount to an attack on sovereignty (think of refusing to accept the Mormon woman's insistence that she knows what she's doing, she really wants to become the third wife of the man she is about to marry, and really, who are you to interfere with her most personal choices?). If this—that sovereignty is not compromised by the relevant preferences being adaptive—is small comfort, perhaps this goes to show that there's something superficial about the sovereignty concern.<sup>46</sup>

In terms of nonalienation, one natural move here would be to emphasize internal tensions within the agent, as when a specific superficial preference of hers is in tension with her deeper commitments. Even though these notions—of superficiality and depth, for instance, and of tension—need unpacking, there's a fairly clear intuitive sense in which a preference that coheres with my deep commitments—those that are fairly stable over time, that serve to give structure to my life, that are expressions of what I find valuable and important in my life, and so on—is more fully autonomous than a preference that doesn't. One natural and influential way of filling in the details here—the one I'll focus on (though everything I say about it will generalize to others, I think)—comes from Frankfurt (1971). If one has a preference or a desire that one doesn't endorse, or indeed that one wants not to have or that one wants not to lead one to action, then that preference—and acting on it—may not be autonomous. At the very least, it will not be as autonomous as acting on a preference or a desire that resonates throughout one's hierarchy of higher-order desires. This may be thought of as an account of nonalienation.<sup>47</sup> And it applies, I'm sure, to some cases of adaptive preferences. But as a full account of the autonomy deficit of adaptive preferences this won't do, for the following two reasons. First, many adaptive preferences

---

autonomy. See Oshana (2006: chap. 1). For a similar distinction, see Meyers's (1987: 624) distinction between programmatic and episodic autonomy. Of course, we can expect to see relations between local and global autonomy—presumably, more locally autonomous preferences and choices contribute to a more autonomous life. So perhaps I get closer here to Christman's (2009: 135–36) characterization of his project as a kind of intermediate.

46. That there may be something superficial about it does not mean, of course, that sovereignty is not extremely important, in private as well as political settings. I return to this in the final section.

47. See my "Hypothetical Consent" (Enoch 2017a: 25).

do not fit this pattern. The music lover may prefer (now) philosophy over music, and she may also prefer so preferring, and prefer that this preference (rather than other ones) lead her to action, and so on.<sup>48</sup> The fact that this preference of hers resonates doesn't make it less paradigmatically adaptive. And second, think of the woman who not only prefers to malnourish herself in order to allow her husband to eat much more than he needs but also thoroughly identifies with this preference; assume that nowhere to be found in her hierarchy of ever higher-order desires is anything remotely resembling a qualification, or a rebellious desire against this first-order desire and those that come along with it. This depressing picture, it seems to me, is one where the woman is less, not more, autonomous. But there's no disharmony in her hierarchical structure of desires. So such harmony does not suffice for autonomy.<sup>49</sup>

Such cases—deep oppression cases—count equally against any subjectivist account, any account that fully grounds autonomy in something internal to the relevant agent. For instance, Christman's partly historical, entirely subjective account (2009: 155) in terms of the agent's critical reflection also has to declare the woman—once *sufficiently* oppressed—fully autonomous.<sup>50</sup>

48. For some adaptive preferences, the hierarchical structure may be interestingly different: even when the fox still wants the grapes, he can already have the second-order desire to stop wanting them, and he may identify with that second-order desire. Indeed, this second-order desire may play a causal role in bringing about the change in first-order desire, a change that, though adaptive, may restore harmony.

49. This is a common criticism of Frankfurt. See, for instance, Westlund (2003: 484); Stoljar (2015: 22).

Here's one—even more depressing—way in which Frankfurt can respond. If there really is no disharmony in the woman's structure of desires, then there is no self there that objects or indeed can object to the self-starvation. So there's no way in which the self-starving is nonautonomous. It may be bad or wrong for any number of other reasons. But it is not less autonomous than any other action ever is. This picture is especially depressing because according to it this case marks oppression's absolute victory.

I take it this is an unpleasant bullet to bite, so we should resist this line unless we have no other options. In the text I am about to suggest another option.

50. Christman (2009: 161–62) gets close to admitting this point. For the explicit (surprising) claim that oppression is consistent with autonomy, see also Christman (2004: 156–57). For a similar bullet-biting coming from Killmister (regarding her own, multi-dimensional account of autonomy), see Killmister (2018: 117–18). For a general criticism of such results—very close to the one in the text—see Westlund (2003: 503). Westlund herself thinks that her emphasis on the social notion of responsibility relevant to autonomy introduces external constraints, and renders her account immune to deep oppression cases. For reasons I can't get into here, I don't think this move succeeds.

This observation pulls in the direction of objectivist, content-based accounts of the autonomy deficit in (some) adaptive preferences. For it seems that the problematic feature of the woman's preference (in the circumstances, for self-starvation) is precisely its content. Such malnutrition is so clearly bad for her, and so unfair, you may think, that any preference *for that* is nonautonomous. What grounds its being nonautonomous is not anything about the structure of preferences of which it is a part, or about its history (a point I return to shortly), but rather its content.<sup>51</sup>

You may be worried about this suggestion if you think that moral neutrality is an important desideratum here.<sup>52</sup> The hierarchical model is morally neutral—whether a specific preference satisfies it may be decided independently of any specific moral view, and perhaps for this reason the hierarchical account may have some political advantages. The content-based account cannot, of course, claim moral neutrality: it grounds the nonautonomous nature of some preferences in the *badness* or *wrongness* (or perhaps *depravity*) of their content. But I don't think of moral neutrality here as a desideratum—in some political contexts it may be, but we're not doing politics just yet.<sup>53</sup> In trying to understand how a preference may show an autonomy deficit, I don't see why we should expect or want to remain morally neutral (and in what follows, I won't).

Still, I don't think that a content-based account is the way to go, for the following three reasons. First, think about a case in which someone desires or prefers the bad, and by utilizing unjust means we intervene and succeed in shifting their preferences for the better. In terms of content, the new preferences are just fine. Still, it's intuitively clear that they suffer from an autonomy deficit. So focusing on the moral value of the content of the relevant preference cannot be enough.

51. For examples of such accounts—sometimes labeled substantive accounts of autonomy—see Hampton (1993: 152), and Superson (2005), though Superson insists (122) that her normative constraints on content are constraints of rationality, not of morality. For many more references, see Killmister (2018: 85n30).

52. See, for instance, Dworkin (1988: 8), Christman (2015: 8; 2004).

53. Christman (2009) insists on a moral neutrality constraint here, partly because he ties the understanding of autonomy more closely to politics than I do, and partly because—in politics—he assumes an extreme kind of (perhaps Rawlsian) view that pluralism requires neutrality (Christman 2009: 5–6, 14, 139, 161). I reject such thoughts about pluralism and neutrality, in the context of rejecting public reason theories, in my “Against Public Reason” (Enoch 2015) (and elsewhere).



Second, it seems to me that for any preference-content, we can describe a possible context—historical and otherwise—that renders that preference autonomous.<sup>54</sup> Think, for instance, of the woman who chooses to malnourish herself in order to benefit her husband, but does this in a context where such a decision is neither common nor expected, and where there are no gender norms that push her in that direction; perhaps she does this out of the purest love (which she fully identifies with). This may not be an ideal situation, of course, but I don't see that it's plausibly considered a failure of autonomy. And if you're not happy with the example, I'm sure you can find others. The main point is simply that for any preference-content, circumstances may be thought of where the preference is not plausibly considered nonautonomous. And if so, no content-based account of the autonomy deficit of adaptive preferences can succeed.

Third, and relatedly, I don't think there's intuitive pressure pulling in the opposite direction. The force of the relevant examples—like the self-starving woman—heavily depends on the social and historical context. It is no coincidence, after all, that the environment in which the woman's preference is shaped is strongly patriarchal and oppressive. So perhaps the thing to look at is not the content of the preference, but its causal history.<sup>55</sup>

What is it about the causal history of the woman's preference that makes it nonautonomous? It can't be the mere fact that her preference (for self-starvation) *is* caused by external factors—that it is not, as it were, fully self-created—that makes it less than fully autonomous. Presumably, none of our preferences are fully self-created, and yet some are autonomous.<sup>56</sup> What we need, then, is a distinction among causal histories of preferences between those that undermine (at least partly) the autonomy credentials of preferences and those that don't.

One influential family of attempts to distinguish the causal histories that undermine autonomy from those that don't attempts to do so in descriptive, nonmoral terms.<sup>57</sup> Coercion, say, undermines autonomy.

54. "There are no *a priori* truths about the content of an autonomous person's values" (Dworkin 1988: 29). Also see Stoljar 2018: 234, 237, and the references therein, and Killmister 2018: 9.

55. For the intuition that nothing about a person's time slice will suffice to distinguish between cases with and without autonomy, see Christman 2009: 137.

56. An often-made point. See, for instance, Oshana 1998: 84.

57. One example is Christman, with his original historical view now revised (2009:

Perhaps so does manipulation.<sup>58</sup> Perhaps the list of autonomy-undermining historical factors is longer, though probably not by much. And then, a decision, or choice, or perhaps preference is autonomous if it was formed in a way free of the causal influences of these autonomy-undermining factors. And if this list of autonomy-undermining factors can be accounted for in nonmoral terms, then this kind of historical account can maintain the kind of moral neutrality the content-based accounts cannot. Or perhaps we can go not for a list, but for a more unified, perhaps subjectivist, still descriptive characterization—say, in terms of whether the agent, had she reflected on the historical causes of her preferences, would have endorsed them (Christman 1991).

Still, such nonmoral, historical accounts (often referred to in the literature as *procedural* accounts of autonomy) face insurmountable problems.<sup>59</sup> Deep oppression cases seem to refute the more unified, subjectivist versions of such an account, and perhaps others as well (it's not clear in what sense, after all, the deeply oppressed are necessarily coerced or manipulated) (Stoljar 2015: 22). The objective list versions also face difficulties in filling in the details. One problem is that I don't think the prospects of understanding coercion (or delineating the coercive cases that undermine autonomy) in nonmoral terms are promising (Nozick 1969), and the prospects for doing so vis-à-vis manipulation seem even less promising (Buss 2005). Finally, while it would be good to have a list of autonomy-undermining factors, it would be much better to have a unifying rationale for the items on this list—what, if anything, explains why those factors undermine autonomy and others don't? At the very least, an account that can supply such a unifying explanation is for that reason pro tanto better than one that cannot. So the prospects for a nonmoral, historical or procedural account do not seem promising.<sup>60</sup>

---

chap. 7). The revised view still attempts to do everything in morally neutral terms, but is less paradigmatically historical.

58. But see Buss (2005).

59. I think this applies also to Dworkin's mature account (1988: 20), which combines historical elements with hierarchical ones, but characterizes all of them in morally neutral ways.

60. Weberman (1997) puts forward what he takes to be a historical nonmoral account. But his account utilizes such concepts as *social domination*, and it's hard to see how they in turn can be understood without a significant moral or at least normative component.

Without pretending to have ruled out all other options, let me put forward a positive proposal, conceding that more needs to be done to support it.

My suggestion, then, is that an important class of cases of nonautonomous preferences is those that were shaped (in the appropriate way) under the causal influence of *unjust* conditions, conditions that violate the rights or entitlements of the relevant agent.<sup>61</sup> You see how this works, most naturally, in the case of the self-starving woman: her relevant preferences were not merely shaped—like all preferences—under the causal influence of all sorts of factors external to her. Crucially, they were shaped under the causal influence of a social order that routinely violates her rights, that wrongs her. Preferences shaped in this way are nonautonomous, even if endorsed. Indeed, if they are endorsed, this may just indicate the depth of the oppression, not its absence. The causal role played by injustice here explains why it is that these preferences (or consent or choice based on them) do not manifest the value of autonomy (as non-alienation), for it shows an important sense in which these preference are not truly the agent's. The fox is different, though. His preference (against having the grapes) was causally affected by the infeasibility of getting them, and if this is, as it seems, a reason of the wrong kind for a preference then this is already something problematic about that preference. But it was not shaped by injustice or violation of the fox's rights, and so it is not in the relevant sense nonautonomous. The situation would have been different, on my suggested criterion, had the grapes been rendered unattainable by an act of injustice: if, for instance, the fox was entitled to the grapes, which were quite achievable, and Aesop took them away, and rendered them unachievable (he needed the story), and if then the fox's preference changed accordingly and he came to prefer not having the grapes, *then* this preference (or change in preference) would have been nonautonomous.

On this suggestion, then, an important feature of some preferences is the moral status of the historical circumstances that shaped them causally. And while this is clearly a historical criterion, it is also a moral

61. This suggestion is inspired by Nozick's (1969) introduction of a moral baseline, in the context of discussing coercion. For discussion, and for the claim that coercion is "thoroughly moralized," see Berman (2002: 45). Not much in what follows will depend on the precise way of delineating the unjust within the immoral. In particular, I do not assume a specific view of injustice as necessarily political or necessarily political in some way. (I thank Sarah Stroud, David Plunkett, and Becca Rothfeld for related points.)

one.<sup>62</sup> Still, despite not being morally neutral, it retains three advantages over the other nonneutral kind of account, the content-based one: it is not subject to the counterexamples where unjust means are used to shift an agent's preferences from the bad to the good; for any preference-content, it allows for the possibility of an autonomous preference with that content (if it were shaped under conditions that did not include an injustice of the relevant kind); and it gives pride of place to the causal history of the preference, as seems appropriate.

The expression "in the appropriate way" calls for (perhaps skeptical) discussion, and I get to it in a couple of paragraphs. First, though, I want to note two important clarifications about alienation, and about the relation, on the suggested account, between the value of nonalienation and the causal influence of past injustices.<sup>63</sup> First, while I insist that some preferences—in our context, those that have been shaped by injustice—fail to manifest the value of nonalienation, I do not claim that in those cases *the agent is alienated from her preferences*. A natural—perhaps the only natural—understanding of such a phrase as "agent A is alienated from her preference P" is about some internal tension within the agent, as when A has a higher-order preference against having P, or some such. And while it's true that if the agent is in such a way subjectively alienated from her preference P, this means that P (and acting on it) doesn't fully manifest the value of autonomy (and so suffers from an autonomy deficit), this is not the only kind of case in which this is so. In deep oppression cases, as noted above, there may be no internal tension, and so perhaps it would be odd to say that the self-starving woman is alienated from her relevant preferences. But still, these preferences of hers suffer from an

62. The account Terlazzo (2016) puts forward—and calls "an indirectly substantive account"—applies substantive, moral evaluation not to the content of the relevant preference, but to the process of shaping it. The details differ—she emphasizes the value of the options at the time of the shaping of the preference, I emphasize the causal relevance of injustice, and while there may be relations between the two suggestions, they are not very close. And there are other details in Terlazzo's discussion (both substantive and methodological) that I don't accept, including her argument for her indirectly substantive account. Still, in putting forward a historical-moral account, I am following in her footsteps.

For a comment suggesting something along these lines, see Cudd (2006: 183), referring to *coercive* social circumstances in the history of the preference formation. Walker (1995: 459) speaks of adaptive preferences as those formed in response to injustice, but he does so as a matter of stipulative definition (for the term "adaptive preference") rather than as a substantive claim.

63. I thank an anonymous referee for pressing me on these points.

autonomy deficit. Of course, it's not as if the thing to say about her is that these preferences of hers are fine, as far as alienation goes, because, after all, it's not the case that she's (subjectively) alienated from them. Rather, the thing to do is to drop talk of whether or not she's alienated from her relevant preferences, because such talk seems to presuppose (or at least to suggest) the kind of subjectivist account of autonomy we've already seen sufficient reason to reject. A less misleading way of talking here would be directly in terms of the question whether the relevant preferences manifest the value of autonomy (as nonalienation).

Which brings me to the second, to an extent methodological, point. The account here suggested is not meant as an account of what autonomy *is*. One reason for that is that I fully acknowledge that my account, even if successful, is only a partial one: preferences (and choices and decisions and so on) may fail to manifest the value of autonomy for more than one reason. Having been shaped by injustice is just one of the ways of grounding an autonomy deficit. A preference may be less than fully autonomous because, say, the agent is alienated from it in the subjective sense from the previous paragraph, or perhaps because it is based on a false belief, or—and this is an important alternative, certainly in oppressions cases—because it has been shaped by injustice. (I return to this below, in showing how my account is consistent with a hybrid account of autonomy deficits.) Clearly, then, I do not claim that not having been shaped by injustice is what autonomy *consists in*. Indeed—and this is the second reason why I'm not putting forward my account as an account of what autonomy is—I don't think that such an account is either possible or necessary. We have pretheoretical intuitions that seem to capture the intended value of autonomy as nonalienation—the intuitions captured by the examples and the metaphors I used when introducing such autonomy talk above, in section 1. Hopefully, those helped you see, in an attractive light, the kind of value I—and many others—are after. What follows—and in particular, the suggested account in terms of the causal role of a past injustice—is an attempt at finding a necessary condition for the presence of the value of autonomy, intuitively characterized. In this respect, I think, the inquiry here is in line with most ethical inquiry into conditions necessary or sufficient (or both) for the presence of given values.<sup>64</sup>

With all of this in mind, then, why think that the condition highlighted above *is* necessary for autonomy? That is, why think that having

64. Nor is it special in our specific context—all the accounts surveyed and rejected above are offered, I think, in the same methodological spirit.

been shaped by an injustice undercuts the autonomy value of a preference and of acting on it? We should return to the underlying intuitions and metaphors. If a preference was shaped in the appropriate way by an injustice, and the agent proceeds to act on it (without having made it her own; see below), it seems clear that she is not here occupying the role of the author of her life story. Rather, she's more like a pawn in someone else's game.<sup>65</sup> She is not shaping her own life, but rather allows her life to be shaped (indirectly, via her preferences being shaped) by her oppressors.<sup>66</sup> Even though, then, there are other ways in which an autonomy deficit may emerge, and even though past injustices are not relevant to what autonomy consists in, still, the condition highlighted here—in terms of the causal role of injustice—makes some preferences less than fully autonomous.

With these clarifications in mind, let's return to some complications that may be grouped together under the phrase "in the appropriate way." For not any way in which a preference is caused by injustice suffices for an autonomy deficit. First, typically what's needed for an agent's preference not to be autonomous is not merely that the causally relevant circumstances included an injustice. It has to be an injustice *toward that agent*.<sup>67</sup> Second, it's not sufficient for the causal history of the preference to include the injustice. Rather, the shaping of the preference

65. Paraphrasing Kane (2005: chap. 1).

66. Shouldn't we get rid of these metaphors, and supply more details—perhaps an analysis—of what it is that shaping one's life or authoring one's life story comes to? Shouldn't we then use this analysis to argue for the claim that such autonomy is of value, and furthermore, to determine whether some condition (like the historical-moral one in the text) is necessary or sufficient for the presence of this value? Let me concede immediately both that it would be great to have such an account and that I don't. But this is not a serious flaw, I think, for the following reasons. First, justifications come to an end somewhere, and this—the value of autonomy, intuitively and imprecisely characterized—seems like a good candidate for where justifications come to an end. And second, if there is much more to say here, I am willing to gamble on the content of this future story that it will support what I'm saying here in the text. I am willing to make this gamble precisely because of the strength of the underlying intuitions and the fact that the point in the text is so loyal to them. (If such a future story does not support what I'm saying here, though, and if it's sufficiently plausible, then my account in this article will have to be rejected.)

67. This also means that all that's needed here is *pro-tanto* injustice. If the pro-tanto injustice toward the agent (that plays an appropriate causal role in shaping his preferences) is outweighed by other considerations, this may justify the relevant action, but it does not make the relevant preferences autonomous. (I thank Simon Rippon for drawing my attention to this point.)

has to be sensitive to the unjust circumstances being, well, unjust. Think of the following example:<sup>68</sup> if someone kidnaps me and holds me in a room for many years; if there happens to be a piano in that room, and not much else; if I develop a preference for playing the piano; and if after being released I retain my preference for playing the piano—then it seems like the injustice of the kidnapping, though a part of the causal history of the piano playing preference, does not render it less than fully autonomous. Intuitively, this is because the causal relation between the injustice and the preference is not of the appropriate kind, perhaps because the fact that the stay in that confined space *was* an injustice is entirely tangential to the piano-playing story. But I don't want to pretend that this story makes things here fully transparent. Here as elsewhere, it's not clear how to deal with "deviant" causal chains. I'm going to have to settle, then, on the combination of the "appropriate way" terminology, the intuitions about the involvement of the injustice, and the fact that often we know a deviant causal chain when we see it. In the cases of preferences formed under oppression, for instance, the causal chain does not seem to be deviant.

I have to address four more complications. First, while I've been talking as if preferences are either autonomous or not, really this is a matter of degree,<sup>69</sup> and my suggestion can supply the needed gradability, at least along two dimensions: the relevant injustice may be more or less serious, and its causal relevance (in shaping the preference) may be more or less dominant compared to other parts of the preference's causal history. Other things being equal, then, the more serious the injustice, and the more dominant it is in the causal history of the relevant preference, the less autonomous the preference is.

Second, the injustice cannot itself—on pain of vicious circularity—be characterized in terms of this very assault on the agent's autonomy. But this doesn't mean it can't be characterized in terms of autonomy at all. For instance, a violation of autonomy at an earlier point in time may—if it causally shapes the relevant preferences in the appropriate

68. For which I'm grateful to Tricia Magalotti. For a very similar example, see Christman 2009: 157–58.

69. See, for instance, Meyers 1987: 625; Oshana 1998: 93; and Stoljar 2014. For Killmister (2018) it's a central claim that autonomy comes in degrees along different (four) dimensions.

way—explain why those preferences are less than fully autonomous at a later point in time.<sup>70</sup>

Third, note that the suggested account—in terms of the role of an injustice in the causal history of the relevant preference—is of a necessary condition for autonomy, not necessarily of a sufficient one (the necessity claim too is qualified below). And this means that I can remain open to the possibility of a hybrid account. Perhaps a number of features may render a preference less than fully autonomous. Perhaps, for instance, a fully autonomous preference is only one that is *neither* subject to Frankfurtian disharmony (in one's preference hierarchy) *nor* one that was causally shaped by injustice. Perhaps this is a helpful way to think, for instance, about the case of someone who—under the influence of poverty and hardship—has come to settle for very little, not to prefer more, indeed perhaps to prefer not to have more. If the causally relevant hardships were also an injustice, then these preferences of his are nonautonomous. Even if they weren't, if his higher-order desires rebel against them, if he doesn't endorse them but is nevertheless moved by them, this is not fully autonomous either. But if there was no injustice, and there is no disharmony, then these preferences—nonideal as they may be in other respects—may have no autonomy deficit whatsoever. And there may be yet other necessary conditions for a preference being autonomous.<sup>71</sup>

Fourth, there is a dynamic aspect to the autonomy status of a preference. It seems possible—and it seems important that it is possible—for an agent to make a preference that suffers from an autonomy deficit his or her own, to vindicate its autonomy status by retroactively and autonomously endorsing it.<sup>72</sup> In such cases, arguably, what used to be a non-autonomous (or less autonomous) adaptive preference becomes (more) autonomous because of such an endorsement. In this way, then, the making-one's-own phenomenon is an exception to the claim that the

70. I thank Tom Kohavi and Massimo Renzo for making me see this point. And see Killmister 2018: 117, for a related point.

71. For one thing, I haven't said anything about either the competence conditions necessary for autonomy (a point Mollie Gerver emphasized in discussion), or the relevant information conditions (as, clearly, false beliefs, and sometimes even just ignorance, can undermine the value of autonomy and the full force of consent—a point made by an anonymous referee). Also, perhaps these other conditions are where more subjectivist intuitions may be accommodated, and where there's room on my account for some of the multiple dimensions Killmister (2018) emphasizes.

72. I thank Stéphane Lemaire, Preston Werner, and Dani Attas for making me see this; and see related comments by Terlazzo (2016: 225).



moral-historical account identifies even just a necessary condition for autonomy. It's an interesting question what is needed for the making-one's-own move to itself be autonomous and so for it to remedy the preference's initial autonomy deficit: Perhaps, for instance, the agent has to understand the initial autonomy deficit and its grounds, and then—with her eyes wide open, as it were—endorse it. Or perhaps there are other conditions—on top of this one, or instead of it. Be that as it may, I want to fully acknowledge both that the phenomenon of making a nonautonomous preference one's own is very real, and that theorizing it requires many more details than I've supplied here. (I do a little more in a long footnote.)<sup>73</sup> But I do want to explicitly leave room in my account of nonautonomous preferences for this important phenomenon.

On this suggested account, then, the full story explaining why a preference does or does not suffer from an autonomy deficit (and how

73. Here, then, are some relevant points:

- The point is often made that the condition for an autonomous action or preference cannot be too intellectual. I agree with this point in general, but I don't think it applies to making-one's-own cases. Given the original injustice and the autonomy deficit it creates, it's quite possible that serious reflection is needed to render the relevant preference autonomous. Indeed, perhaps being willing to hold one's commitments up for inspection (Railton's [1984: 147] phrase) is sometimes even required.  
(Notice a terminological problem here: I think of nonalienation as the relevant kind of autonomy here. Railton thinks of holding one's commitments up for inspection as possibly alienating, yet still sometimes required.)
- Similarly, Christman (2009: 143–44) emphasizes that requiring active endorsement as necessary for autonomy would be too demanding. In general, I agree. But requiring something active for making a nonautonomous preference autonomous does not sound too demanding.
- It seems plausible that for an agent to make a preference her own it's necessary that the motivations underlying this very process do not themselves suffer from an autonomy deficit. So some kind of a recursive condition seems needed here as well.
- An interesting case (for which I thank Hallie Liberto) is of a preference or action motivated by a false moral belief, but where the false moral belief is caused without an injustice (perhaps by reading it in a mistaken book). I think this need not be a case of a failure of autonomy, and would apply the general things we have to say regarding preferences under false beliefs, but I'm not sure this is satisfactory here.
- Another possible way of thinking about the making-one's-own phenomenon is as helping to salvage a more global sense in which one's life is autonomous, even if it fails to render the specific preference fully autonomous. (I thank Avishay Ben Sasson Gordis for this suggestion.)

serious the deficit is, if there is one) will need to incorporate a full substantive theory of injustice. This may seem like a shortcoming, but in fact I don't think that it is. First, while the *full* story will have to include all the considerations that may render the conditions playing an appropriate causal role in the shaping of preferences unjust, we rarely need the full story. Very often, we will be able to rely on paradigmatic injustices in order to know that a preference is less than fully autonomous, in a way that will not require invoking a specific, controversial theory of justice. Second, the relevance of considerations of justice seems upon reflection to be entirely unsurprising. In trying to determine whether a specific decision or a specific preference is in the relevant sense the agent's own—whether it is in line with the value of autonomy as nonalienation, whether it has the normative implications (discussed in the next section) of autonomous decisions—why think that considerations of justice in general should *not* be relevant? Still, this gives rise to a related suspicion,<sup>74</sup> for now the danger of counterexamples looms large. Plug in a fairly expansive theory of injustice, and you'll get the result that very few of our preferences are autonomous. I think that the worry may be exaggerated, but that at the end of the day it needs to be tackled head on. Recall that whether a preference is autonomous is, on my account (and any other plausible one), a matter of degree. With this in mind, the suspicion should be restated, presumably, as the thought that on some substantive theories of injustice, it will follow that many of our preferences are less than fully autonomous. This is a sad thought, perhaps, but not an implausible one. So the worry is exaggerated. Still, that many of our preferences are less—perhaps far less—than fully autonomous remains a worry. But, I want to suggest, this is a worry about politics (and perhaps the human condition), not about my theory. Indeed, it's an advantage of the theory that it can be productively used in diagnosing such autonomy deficits, and perhaps also—further down the route—in fighting them.

#### 4.2. *Why It Matters Whether Adaptive Preferences Are Less Than Fully Autonomous*

“Autonomy” is not an empty honorific bestowed on the preferences (or decisions, or choices, or lives, or people, or states, or . . . ) close to our heart. Proclaiming a preference nonautonomous has normative implications, in ethics and in political philosophy, and these implications are

74. For which I thank an anonymous referee.

what make this classification of preferences into more or less autonomous important, and—therefore—they are also what determine to an extent what the word in our context means.<sup>75</sup> But because whether or not a preference is autonomous may matter in more than one way, we must again draw distinctions. And we must then check how adaptive preferences (of different kinds, perhaps) do vis-à-vis the different implications of their purported autonomy status.<sup>76</sup>

#### 4.2.1. Responsibility and Agency

I have—already at the outset—set aside the kind of autonomy talk that is more closely related to free will, agency, and perhaps the relevant Kantian inspiration. But I want to briefly revisit this issue just in order to emphasize the following: even if we find out that an action is rendered less than fully autonomous, perhaps because motivated by an adaptive preference, in the senses below—in terms of the extent to which it manifests the value of autonomy, or in terms of the justification of noninterference—nothing yet follows about the agency or moral responsibility of the person whose action it is. These are distinct questions, and this last set of questions—about agency and responsibility—is just not one I will be discussing. What the conditions necessary for moral responsibility are is, of course, a complicated question, and while it may involve matters that are close to the ones I will be discussing in the rest of this section, it's also possible that it will not, or that it will involve so much by way of other factors that these ones will be crowded out. Similarly, of course, for agency, and indeed, for the complex relations between agency and responsibility.

It's important to emphasize this in our context, because some of the relevant discussion in the literature is being held, it seems to me, with the thought—sometimes explicit, often not—that there is something offensive about proclaiming a preference not fully autonomous. Perhaps

75. For similar methodology, see Dworkin 1988: 6.

76. There's a sense, then, in which the term "autonomy" may at the end of the day drop out of the picture. We are ultimately interested in the relation between the feature that may make a preference less than fully autonomous, and the different implications of its being less than fully autonomous. It will sometimes be possible to proceed from the former directly to the latter, without making a stop at autonomy talk.

This is true, but unhelpful. First, autonomy talk here does do some unifying work—it places different (but related) concerns in the context of the very natural thoughts we started with of being an author of one's life story, and the like. And second, the word "autonomy" may indeed be dispensable, but only in the way in which every word is. We care about what we say, not about how we say it.

by doing so—especially in the case of preferences shaped under oppression—we add insult to injury, delivering the message that the person is not fully an agent.<sup>77</sup> Far from taking a step in understanding oppression and fighting it, it may seem like we’re reinforcing it. This is not so, though, for two reasons. The first is the one just highlighted. The discussion that follows is not about agency. Nothing here calls to question the agency—in the fullest sense of this term—of the Mormon woman, say, or indeed (in a less important way) of the fox. And second, no one is *proclaiming* any preference nonautonomous, or, for that matter, anything at all. Ours is a philosophical discussion, and we’re looking for the true answers to the questions we’re asking ourselves. What—if anything—we should then do with them, including whether we should proclaim them and how, is a matter for another discussion, the more clearly political one to which I return below. If shouting our (true) conclusions from the rooftops will make all hell break loose, this is a very good reason not to shout them from the rooftops, but it is not any reason at all not to believe them.

#### 4.2.2. How Well One’s Life Goes

I can think of two ways in which the autonomy deficit of adaptive preferences may make a difference to well-being. The main one I want to discuss here has to do with the value of autonomy. But first, I want to talk about preference satisfaction.

If you think that well-being is at least partly constituted by preference satisfaction, you may want to ask whether it matters, for its satisfaction’s contribution to well-being, that a preference is adaptive, and so that it’s perhaps less than fully autonomous in the way described above. In particular, does the fact that a preference was causally shaped by injustice affect whether, or to what extent, satisfying it will contribute to the agent’s well-being?<sup>78</sup> In the opposite direction—does the extent to which getting

77. This is a central concern for Begon (2015). See also Stoljar 2018: 233. Stoljar also distinguishes autonomy from having an authentic agential perspective—a distinction that’s close to the kind of thing I’m insisting on in the text. In this context see also Terlazzo 2016: 208–9, and the references therein.

78. Relatedly, assuming preferences have some significant role to play in an agent’s deliberation—because they are normatively relevant, either for well-being or in some other way—should that deliberative role be sensitive to the relevant preference being adaptive? Bruckner (2009) answers in the negative. But I think that things here are more complicated: if the agent doesn’t know the relevant preference is adaptive, he (subjectively) ought to give it the standard role preferences should get in deliberation, of course,

what you want is (intrinsically) good for you depend on why you want what you want? The full answer here will depend on the motivations for a preference-based view, and on the nature of the relevant preferences. On a simple view of preferences as nonrational states, perhaps analogous to itches, nothing in their history seems to matter. An itch, it seems to me, has the normative implications of an itch—a reason to scratch, perhaps—regardless of its history, or of anything else about it (perhaps especially weird cases aside). If preferences are like this, then once they're there, they're there, with all their normative implications in place. Of course, this is a *very* crude view of preferences. And the more complex and nuanced one's view of the nature of (at least some) preferences, the more plausible will it be to claim that the preference's history matters, and so in particular, that it matters whether it's autonomous and whether it's adaptive. For instance, if one thinks of (at least some) preferences as *already* responses to reasons, then much of the discussion above becomes more clearly relevant. The problem, though, is that then a preference-based theory of well-being becomes less plausible. After all, if one is happy to include objective normative elements in one's understanding of preference, it's no longer clear what advantage one has over a more directly objectivist view of well-being. So while I think there's room in logical space for a preference-based view that accords significance to the distinction between adaptive and nonadaptive preferences, it doesn't seem to me likely that it will be well motivated.<sup>79</sup>

But now think of the view according to which living an autonomous life is pro-tanto intrinsically good, that is, that a life in which one is a part-author of one's life story is, other things being equal, a better

---

but this says nothing about adaptive preferences—it's just how the subjective ought works. And a situation in which an agent has an adaptive preference, recognizes that it is adaptive, and continues to have it is sufficiently nonstandard to undermine any confidence in judgments about it.

79. A fuller discussion of the point in this paragraph would have to address also pluralist and hybrid accounts of well-being—accounts that accord some weight to preferences or other subjective states, but also to more objective ones.

And here's a related case (which I cannot discuss in detail here): suppose that preferences are reason-giving also for other people—we sometimes have a reason to give people what they want, or to help them get it, and so on. Are such reasons sensitive to whether the preference (or its absence) is adaptive? If we can very easily get the fox the grapes, do we have a preference-based reason to do it, even now that he no longer wants them? We may have other reasons to do so—perhaps he will still enjoy them. But it seems to me that we no longer have a preference-based reason to get him the grapes. It's not clear to me whether this conclusion generalizes.

life. The value of autonomy thus understood is manifested when, say, you make an important choice, perhaps among an adequate range of options—a choice that determines something significant about what your professional life looks like, or your personal life. It is the intuitive plausibility of this value that I highlighted in section 1. And now we can ask the following question: when a preference shaping a choice or a decision is adaptive, does this affect the extent to which it manifests the value of autonomy thus understood?

Keep in front of your mind's eye, then, the intuitive force of the value of being the author of one's life story, and think about some of the examples. The woman who prefers to malnourish herself so that her husband and male children can have more food than they need, where this preference was shaped by injustice—when she chooses yet again to deprive herself of some food, does that choice of hers manifest the value of being the author of one's life? Is there more value in her life, as it were, even *pro tanto*, in virtue of making this choice? Is a life that consists of many central choices that are like this one a paradigmatically autonomous life? I submit that the answer to all these questions is a clear “no.” And this, it seems to me, is the most important way in which it matters that a preference is adaptive (in the specific, injustice-involving way)—it deprives acting on it from the positive autonomy value it otherwise would have had.

I hope you're with me on the intuition in this case. But in order to boost its plausibility, it may be helpful to clear some possible misunderstandings.

First, the question at this point is not practical. The question is whether the value of autonomy is manifested in this choice, not what to do about it (I get to this shortly). So a negative answer to our question doesn't entail anything about state intervention or the like.

Second, partly because of that (and echoing a point above), the mere realization that the value of autonomy is not present in this kind of choice is in no way an insult to the person making the choice. The fault rendering that choice nonautonomous in the relevant sense is not in her, but in the situation she finds herself in, perhaps in the others putting her in that situation.<sup>80</sup> Think in this context of Raz's (1986: 374) plausible insistence that the value of autonomy requires not just the relevant competences and the absence of certain obstacles (like coercion), but also an

80. A point emphasized in this context by Terlazzo (2016: 221).

adequate range of options—that is, sufficiently many, sufficiently diverse, sufficiently good options—to choose from. When someone doesn't have available to them such a range of options, their choice will not manifest the value of autonomy. But this in no way reflects badly on them. Indeed, a part of the problem in the woman's case may be precisely that the range of options available to her is far from adequate. Perhaps, to repeat, she rationally opts for the best option available to her given the social complexities. But my point here is that even if she doesn't—even if there are better choices available that she doesn't opt for because of her adaptive, nonautonomous preferences—understanding this fact does not, and certainly need not, amount to an insult. In this respect, the situation is similar to that of the example of the Walmart employee. Even if in his case the relevant preferences are not exactly adaptive, the fact that they are based on manifestly false beliefs, and indeed beliefs that the unjust social order invests resources in propagating and inculcating, means that his political choices do not manifest the value of autonomy, but this need not be his fault—it may very well be others' fault (though sometimes there's enough fault to go around).

Third, of the two autonomy values I distinguished—sovereignty and nonalienation—the one that seems threatened in this kind of case is that of nonalienation. Sovereignty seems more relevant to the topic of noninterference (discussed in the next subsection). When the woman malnourishes herself, this in itself does not seem to be an offense against her sovereignty<sup>81</sup>—indeed, perhaps not allowing her to would be such an offense. But in terms of the value of nonalienation, the problem seems clearer. That value, recall, is the value of living one's life according to one's deep commitments. The problem with the woman's commitment to depriving herself of food is not necessarily its depth; rather, it's that given its causal history, it's not clearly *hers*, it's not a manifestation of self-authorship in the sense in which such authorship is of value.<sup>82</sup> It's a commitment she only has because of her having been the victim of injustice.<sup>83</sup> Living one's life according to the preferences and commitments one only

81. Though not much here is obvious. Perhaps what we want to say is that while there's no offense against her sovereignty here, there's no manifestation of it either.

82. So what the oppressive social order does to the woman, in this case, is not prevent her from leading the life that is hers, but rather make it the case that there is no longer a possible life that is really hers.

83. This way of putting things entails that cases of overdetermined preferences—preferences that are causally overdetermined, caused by injustice but also by something else, such that either of those would have sufficed to cause the preference all on its own—

has in virtue of having been the victim of injustice (in the way described in the last subsection) does not manifest the value of nonalienation.<sup>84</sup>

With these clarifications in place, I hope the point above becomes plausible to the point of irresistibility: the woman's choice to starve herself—whatever its other merits, whatever we should do about it, and without blaming her in any way—does not manifest the value of autonomy. The causal history here makes it the case that the value of autonomy is absent (at least to a large extent). And there's no mystery about this: the appeal of living according to one's own preferences and commitments is lost when we add to the picture the unjust causal history that shaped those preferences and commitments.<sup>85</sup>

#### 4.2.3. Noninterference

Certainly in the context of consent, but not just there, autonomy talk is often invoked not primarily as a talk about the quality of life of the autonomous, but rather as a call for restraint against interference, often by the state. Thus, the Mormon woman from Mill's example may insist that she's not interested now in our view about which values are manifested by which choice of hers; she just wants us out of her marriage. The point is not in the first instance about value at all. It's about noninterference.

The most important thing to note about this autonomy-related line of thought is that it is distinct from the previous one.<sup>86</sup> Thus,

---

may yet be autonomous. This sounds plausible to me. I thank Bar Luzon for pressing me on the relevance of overdetermination.

84. I'm not sure what to think in this context of other adaptive preferences, like the fox's or the music lover's. Does living one's life according to them manifest the value of nonalienation?

85. What if the agent's preference was causally shaped in the appropriate way by an injustice, but the injustice was brought about by the agent herself? This kind of case (for which I thank Hasan Dindjer) may challenge thoughts about self-authorship. I think that for the reasons in the text, the presence of injustice and its causal role suffice to ground, in this kind of case as well, an autonomy deficit. If the agent's involvement in bringing about the injustice was merely causal, I don't think that it matters. Things may be different in cases of self-manipulation, where the agent deliberately brings about autonomy-undermining circumstances, precisely in order not to be autonomous at a later point in time. I think that even in such cases, the agent (at the later time) is not fully autonomous, but things are complicated here.

86. Terlazzo (2016: 217) makes a somewhat similar point, but terminologically differently: instead of distinguishing different ways in which autonomy may matter, she severs the tie between autonomy (in the only sense she recognizes) and coercion.

Christman (2004) doesn't draw the distinction in the text, and is thus led to the surprising claim that oppression is consistent with autonomy (156–57). But perhaps the



acknowledging—as I argued we should—that choices motivated by autonomy-suspicious adaptive preferences fail to manifest the value of autonomy in no way entails that the usual barriers against interference are not in place. For anything thus far said, it may not even be relevant to that question—the considerations relevant to whether interference is justified may be blind to the question whether the relevant preferences and choices are adaptive in this way.

Here's one easy, relevant point: even if the contribution of the value of autonomy to the well-being of the relevant agent does not justify noninterference in such cases (because, as I've argued, that value is not manifested here), there may be *other* reasons justifying noninterference. Again, chief among them are instrumental reasons. Interventions tend to be at most partly successful, and expensive in terms of other values and implications. Such considerations can strongly count against intervention—often conclusively—even regardless of fancier discussions of the value of autonomy. When it comes to such instrumental considerations, the fact that a preference is adaptive may be relevant, or it may not—this will depend, as always with instrumental considerations, on the contingent circumstances.

Here is another, fairly easy, point: although the fact that the value of autonomy is not present in the relevant cases doesn't establish that the state should coercively interfere, this doesn't mean that *no* implications for state action follow.<sup>87</sup> For one thing, perhaps the state has the general, standing obligation to make available to its citizens the conditions of a good life, including those needed for an autonomous life. If so, by the time the woman faces the choice whether or not to starve herself, the state *has already* failed: it's failed to provide the woman with an adequate range of options. Even if the state doesn't have a (conclusive) reason to interfere now that the problematic choice situation arises, the state may very well have reason to try and avoid such situations. This, then, is another way in which the state can and should be responsive to the value of autonomy in the face of problems having to do with adaptive preferences: it

---

difference between Christman and me is not that big after all—a major motivation for his claims here seems to be precisely the kind of dangers I highlight in the text. Perhaps the two of us agree about the need to acknowledge the distance between an oppressive history of a preference and the relevant political implications, and disagree only about where in this gap to place autonomy talk—closer to the historical injustice (as on my account) or to the political implications (as on Christman's).

87. For a similar point, see Walker 1995: 467–68.

should try to prevent the unjust circumstances that, when they causally shape people's preferences, render them less than fully autonomous.

But it's time to reach the hard problem here. Presumably, in many cases there is a not purely instrumental (pro-tanto) reason not to interfere in people's actions and choices. That reason, it seems, has something to do with the need to allow them to be the authors of their own lives. That reason not to interfere has to do with the need to respect them, perhaps by respecting their choices.<sup>88</sup> And so the tough question here is the following: is the fact that a preference is adaptive and suffers an autonomy deficit in the way described relevant to the noninstrumental reason against interference? In other words, are choices motivated by adaptive preferences less worthy of protection from interference simply by virtue of being adaptive?

Recall that we are not talking about the choice of a best available option from a very bad set. Those are not the cases of adaptive preference we've been talking about. And with regard to those cases, it seems very clear that the mere fact that the preference or choice *is* made from among a very bad set does not in any way diminish the weight of the reason not to interfere. The woman for whom society has left only such poor options that bargaining with the patriarchy in all sorts of ways is her best remaining option is entitled, it seems, at least to our not depriving her of what is now her best option.<sup>89</sup> Indeed, in such cases it seems to me—and here I may be departing from Raz<sup>90</sup>—that the value of autonomy may still in a way be present. Compare a scenario in which, in the face of only bad options, you find yourself somehow with the least bad one, to a scenario in which, in those circumstances, you actively opt for the least bad one. There is something of value in the latter that's missing from the former, and it seems to me to be the value of being the author of one's life story—

88. This is *not* about respecting people by respecting their false views. It's about respecting their choices (within perhaps the standard limits) even while not respecting their preferences or views or values. For the claim that respecting people does not amount to respecting their false beliefs (but rather to engaging them), see Raz 1998: 143. For my reservation about this claim, see Enoch 2015: 136.

89. The real world is messy, and sometimes tragic. So there may be cases in which fighting an oppressive system will require making some bad-but-best-available options unavailable for some. What I say here in the text is not intended to rule this out.

90. Raz seems to think that the value of autonomy is only present when a choice is made among valuable options. See Raz 1986: 379, for his claim that a choice between good and evil is not enough. As far as I know, Raz nowhere addresses an example of the kind in the text here.

unfortunately, often the very bad parts of our lives are important parts of our lives' stories.<sup>91</sup> In cases in which the agent opts for the best option available to her, then, the fact that it's the best from a set of bad available options is not a reason to interfere, nor does it even weaken the reasons not to interfere; there may even be an autonomy-based reason not to interfere.

But this, to repeat, is not the case we are most interested in. In the case we are most interested in the agent chooses suboptimally, because her relevant preferences were shaped by injustice. And here, I submit, is the important point: there is no (nonalienation) autonomy-based reason to protect such choices. The choices are not autonomous in the relevant sense, the agent is not being autonomous in making them, the value of autonomy is not present; so the most central, intrinsic barrier to interference is not present here. Sometimes, not interfering in someone's choices—even their wrong ones—may be a way of showing them respect. But you don't show them respect by accepting their choices when those are motivated by preferences they only have because causally implanted in them by their oppressors. Perhaps you respect them precisely by refusing to play along. Or perhaps you can't at that precise context respect them—perhaps this is precisely one of the consequences of the effect injustice had on their will. (You can, of course, show them respect in other contexts.) But the kind of reason you often have not to interfere—respecting people's nonalienation—is not present in the cases dominated by the kind of autonomy-deficit adaptive preference I've been emphasizing.<sup>92</sup>

This may seem like a harsh, even dangerous conclusion. And in a way, it is. (Have you missed the words "false consciousness" in the title?)<sup>93</sup>

91. The discussion of how Jehovah's Witnesses reportedly conducted themselves in Nazi camps seems to be relevant here. See, for instance, Stoljar (2018: 231) and the references there. See also Terlazzo (2016: 216n31).

92. Walker (1995: 462–64) argues to the contrary, but his discussion suffers from a failure to distinguish between sovereignty and nonalienation, and even if restricted to non-alienation, it is not convincing for the reasons given in the text here.

93. This term is not popular these days, not even among, say, feminists who have been emphasizing the political and philosophical significance of what I would call false consciousness (talk of "deformed desires" or "adaptive preferences" sounds perhaps a touch more historically kosher). For exceptions in contexts very close to ours, see Walker (1995: 461), Weberman (1997), and Cudd (2006). Here's Steven Lukes (2005: 13): "To say that such power involves the concealment of people's 'real interests' by 'false consciousness' evokes bad historical memories and can appear both patronizing and presumptuous, but there is, I argue, nothing inherently illiberal or paternalist about these notions, which,

So let me mention the following important qualifications. First, nothing here is dichotomous. Choices (and preferences, and actions) may be more or less autonomous.<sup>94</sup> So the standard, autonomy-related reason against interference may be more or less weighty, not just present or absent. Second, there may, of course, be other reasons not to interfere. And the mere fact that the relevant preferences are adaptive in a way that makes them less than fully autonomous is not a reason *for* interference. It's just that the special protection that some preferences and choices and actions get does not extend to these nonautonomous ones.

If you're not yet convinced, think of interference's nicer cousin, accommodation. The value of autonomy sometimes calls not just for non-interference (from the state, for instance) but for active facilitation or accommodation. This holds, at least sometimes, even in cases where the choice is suboptimal, even just for the agent. This comes up in personal relations as well, as when you should help a friend with his even somewhat ill-conceived project (within constraints). And the explanation, it seems to me, is again in terms of the value of autonomy. By accommodating or facilitating, you are helping your friend write his life story, as you have reason to do. Do you, though, have the same reason to facilitate a decision of his that is nonautonomously shaped by injustices directed at him? I say no. But even more clearly: if there is such a reason, is it *as weighty* as the reason you have to accommodate him when the relevant choice and preference are autonomous? A negative answer seems obvious, and this suffices to show that the fact that a preference is nonautonomous in this way affects the reasons to respect it. And while this doesn't *entail* the stronger point above—namely, that a preference being nonautonomous affects also the reason not to interfere, perhaps to the point of annihilation—it certainly supports it.

So sure, there may be other reasons not to interfere. And the relevant adaptive preferences may be less than fully autonomous to different degrees. But the dangerous conclusion stands: to the extent that an action or a choice is motivated by a nonautonomous, adaptive prefer-

---

suitably refined, remain crucial to understanding the third dimension of power." And here's Gerald Dworkin (1988: 11) expressing a similar thought: "While Marxists have been most vocal in raising the issues of 'false consciousness', and 'true versus false needs', it is important to see that the question is one which a wide range of social theorists must address."

94. Recall the two relevant dimensions: the injustice may be more or less serious, and it may be more or less causally dominant in shaping the relevant preferences.

ence—one causally shaped by injustice—the standard autonomy-based reason for noninterference retreats, perhaps all the way to nonexistence.

## **5. Back to Consent, and Politics**

This was long, and complicated, and a touch abstract. We must now draw the threads together, and return to the topic of consent, and indeed, to politics.

Much of the discussion applies—pretty much as is—to consent. Consent matters because autonomy matters. When—and to the extent that—the consent is motivated by adaptive preferences, there is already something suspicious about it, perhaps because it responds to reasons of the wrong kind, but this doesn't suffice for the consent to fail to manifest the value of autonomy, or for consent not to do its normative magic, for instance by rendering permissible actions that would have been impermissible. If our music lover allows you to take her musical instruments, even if this consent is to a large extent motivated by her adaptive preference to be a philosopher rather than a musician, still if you proceed to take the musical instruments you are not thereby guilty of theft. And respecting the value of autonomy—in those cases as well—includes respecting a space of personal relationships that, as long as they are consensual, should for the most part be protected from interference, by the state as well as by most others. This is why if the state requires that you not take the music lover's musical instruments, it's not just you, but also she, who can complain. Such interference will violate her autonomy.

Not so, though, if the consent is motivated by preferences causally shaped by injustice. If the music lover's preference for a career in philosophy (over one in music) was causally shaped, say, by someone depriving her of music opportunities, or by someone lying to her about her prospects as a musician, then her allowing you to take her musical instruments may not have the full normative force of consent.<sup>95</sup> And anyway, nothing about autonomy will call for respecting that consent. (Though some qualifications will soon emerge.) And similarly in the more important examples. Assuming, as seems likely, that the Mormon woman's consent (to become the man's third wife) was to a large extent causally

95. I think that this is true whether or not it is you who wronged her. Of course, if it isn't you, then you may not know that her consent is not autonomous, and—depending on what we think about the objective and the subjective ought—this may affect whether in taking the musical instruments you fail to act as you ought.

shaped by injustices she was a victim of, her consent does not manifest the value of autonomy, and may not have the normative implications consent—real consent—has. And while interference in the purportedly fully consensual relationship between those involved may be problematic in numerous ways, it will not amount to an assault on the woman’s autonomy,<sup>96</sup> at least understood as nonalienation. To repeat, the mere fact that the consent suffers from an autonomy deficit in this way is in no way a reason for interference. But if there are good enough other reasons for interference, then the usual autonomy-based defenses against interference don’t apply. This is just the dangerous conclusion from above, applied to the case of consent.

But now, politics. For the way from this conclusion to practical, political implications is long and treacherous. By way of a preliminary discussion of some relevant complications, then, let me offer the following concluding comments. (I hope to discuss these complications in future work; this is why there is “Part I” in the title of this article.)

First, let me revisit the cases in which the “adaptive” preference is simply the preference for the best available option, given the infeasibility of better options. So think of the variation of the music lover’s case in which she just prefers a career in philosophy over one in law, and she puts her energy into studying philosophy, partly because she realizes that a career as a great composer is not in the cards for her. In such a case, it is, of course, true that she chose—and ipso facto consented to—philosophy rather than law. But I hope it’s clear there’s no sense in which she consented to not being a great composer—if that option *were* in the cards, and someone deprived her of it, the fact that she (in some sense, voluntarily) puts her energy into philosophy wouldn’t make that deprivation any less wrong. This is more clearly and more importantly true in the bargaining-with-the-patriarchy cases. If the Mormon woman prefers a life as a man’s third wife among the options available to her, but would have preferred other options—like perhaps a monogamous relationship that her community doesn’t make available to her, or a life with no male partner but also without the stigma that such a life comes with in her environment—then her voluntarily marrying that man as his third wife, while it may be thought of as consent for marriage-rather-than-not (in those circumstances) in no way vindicates the deprivation of those other

96. Again—I cannot emphasize this enough—assuming this is not her best available option, which is far from clear in the real world.

options.<sup>97</sup> Voluntarily opting for the best of the bad available options may serve as consent to that-option-rather-than-the-others, but not to that set of options. And while depriving a person in such a situation of their best remaining option would be bad, it would not be better to pretend that the fact that they prefer it to even worse ones vindicates the fact that they only had that bad set of options to choose from.

Second, and now returning to the central cases of not-fully-autonomous adaptive preferences and consent based on them: once we're doing politics, instrumental considerations become extremely important. So even if it's true, for instance, that there is no autonomy-based reason to respect the Mormon woman's consent, still there may be a host of instrumental reasons for the state not to interfere, or at least not to interfere in some specific ways, such as criminalizing the husband's conduct. It goes without saying that a political discussion of specific cases will never be complete without according the appropriate weight to such instrumental considerations.

Third, there are considerations that ride piggyback on autonomy reasons, and that may apply even if the autonomy considerations themselves don't. One such type of consideration is epistemic. Whether a given preference's causal history includes an injustice, and if it does, how dominant that injustice is in its causal history, are sometimes going to be hard things to determine. Some modesty is in place, then, regarding the possibility of assessing them, and perhaps at times norms of epistemic modesty—or of the appropriate way to deal, morally and politically, with uncertainty—may thus be relevant here, and may count against interference even when, in fact, the consent is nonautonomous in the ways specified.<sup>98</sup> The difficulty in distinguishing nonautonomous cases from those discussed two paragraphs ago—of rationally opting for the best option from a bad set of options—adds yet another reason for epistemic modesty. Similarly, and as noted above, the justified delineation of the political protection of a value need not fully match the delineation of the value itself. Partly for instrumental reasons, partly for epistemic ones, and

97. One way of putting the point in the text here is that consent is, at least often, a contrastive notion. One may consent to something-rather-than-another, but not also rather-than-a-third-thing.

98. Weberman (1997) and Terlazzo (2016: 221–22n46) make similar points. But nothing here is obvious. You may think, for instance, that an erring-on-the-side-of-safety norm is reasonable here. This seems sensible, but it is not always clear that the side of safety is that of noninterference.

perhaps partly for yet others,<sup>99</sup> it is sometimes a good idea to politically protect a value with some generality, realizing that in some particular cases such protection may be *too* general. Perhaps autonomy should be protected generally in this way, and perhaps the objection to interference in some nonautonomous adaptive preference cases lies in the periphery of such protection.<sup>100</sup>

Then there are questions of neutrality again. The question whether a given consent is nonautonomous depends, I've been insisting, on whether the preference motivating it is adaptive in the autonomy-undermining way, that is, whether it is the causal upshot of an injustice. And of course, this is the kind of thing that is bound to be controversial in a pluralistic society. We should not—obviously!—expect a consensus on the question whether the preferences motivating the Mormon woman's consent were causally shaped by injustice, because the question whether raising girls and women in that environment constitutes an injustice is a matter of deep controversy, for instance, between non-Mormon liberals and (at least some in) the Mormon community. Some think that—perhaps along Rawlsian, political-liberal thought—in the political context it's important to remain neutral on such deeply controversial matters.<sup>101</sup> Myself, I reject this kind of public-reason neutrality requirement, for reasons that I cannot get into here (see Enoch 2013, 2015, 2017b). But regardless of such deeper-sounding theories, there is that old *modus vivendi* rationale,<sup>102</sup> which takes us back to instrumental considerations. It is often extremely important to reach a *modus vivendi*, and in some such cases it will be required to make serious compromises, including to bracket some considerations of justice (without pretending that we are neutral about them). This may apply to some cases of nonautonomous consent as well.

For all these reasons, then, the dangerous conclusion—that consent based on adaptive, nonautonomous preferences doesn't merit the kind of respect that full consent does, and may not have the full normative

99. I'm thinking here of Raz's claim (1986: 378) that autonomy sometimes counts against the elimination of options even when the remaining range of options is still adequate, because of the symbolic meaning such an act would have.

100. I'm not sure just how significant this point is, because I'm not sure the Mormon woman example should be seen as a peripheral one, I'm afraid.

101. This, to repeat, is a central motivating concern for Christman's (2009: chap. 7) account of autonomy.

102. On Rawls's use of the word "mere" (in "mere *modus vivendi*"), see Williams (2009: 2).



implications of consent—is often attenuated in the political context. But of course, there’s more. Recall the distinction between autonomy as nonalienation and autonomy as sovereignty. The claim that the Mormon woman’s consent does not manifest the value of autonomy, and does not merit the protection that autonomous choices merit, was about nonalienation. But even if everything I said about it is true, it will be *very* hard to deny that her sovereignty is also at stake. Not taking seriously her choice—including her voluntarily entering into marriage as a third wife—does seem to violate her sovereignty. And sovereignty is, of course, *especially* important when it comes to politics (see Enoch 2017a: 35).

So this has to be our next task. We have to understand better the value of autonomy as sovereignty, or the constraints that considerations of sovereignty place (for instance, on interference) even when the relevant consent is nonautonomous in other senses. And we need to understand how sovereignty interacts with other relevant values and considerations, and when it may win the day. I hope to embark on that task in my next article.

But even before we make progress on that crucial issue, it’s not as if nothing has been achieved. At least we know what kind of consideration does *not* call for the protection of choices and consent based on nonautonomous adaptive preferences. Indeed, we know that if *any* autonomy-related value—any consideration having to do with being the author of one’s life story—is to be relevant, it’s just the sovereignty one. Knowing that—that such consent doesn’t manifest the value of autonomy, and that respecting it is not called for by considerations of nonalienation—perhaps we can be at least somewhat more lenient in what is required to outweigh them. And of course, we should try to bring about circumstances in which people do not grant nonautonomous consent, so that nothing about their sovereignty requires that we respect such nonautonomous decisions and choices.

So *of course* consent matters. And *of course* it is often suspicious, from a liberal perspective as from any other. It is suspicious whenever the preferences (and the like) on which it is based are adaptive, and so may respond to reasons of the wrong kind. And when it is causally shaped (in the appropriate way) by injustice, the autonomy rationale for consent’s force is heavily compromised.

## References

- Begon, Jessica. 2015. "What Are Adaptive Preferences? Exclusion and Disability in the Capability Approach." *Journal of Applied Philosophy* 32, no. 3: 241–57.
- Berman, Mitchell N. 2002. "The Normative Functions of Coercion Claims." *Legal Theory* 8, no. 1: 45–89.
- Bruckner, Donald W. 2009. "In Defense of Adaptive Preferences." *Philosophical Studies* 142, no. 3: 307–24.
- Buss, Sarah. 2005. "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints." *Ethics* 115, no. 2: 195–235.
- Buss, Sarah, and Andrea Westlund. 2018. "Personal Autonomy." Stanford Encyclopedia of Philosophy Archive, February 15. plato.stanford.edu/archives/spr2018/entries/personal-autonomy/.
- Christman, John. 1991. "Autonomy and Personal History." *Canadian Journal of Philosophy* 21, no. 1: 1–24.
- Christman, John. 2004. "Relational Autonomy, Liberal Individualism, and the Social Constitution of Selves." *Philosophical Studies* 117, nos. 1–2: 143–64.
- Christman, John. 2009. *The Politics of Persons: Individual Autonomy and Socio-Historical Selves*. Cambridge: Cambridge University Press.
- Christman, John. 2015. "Autonomy in Moral and Political Philosophy." Stanford Encyclopedia of Philosophy Archive, January 9. plato.stanford.edu/archives/spr2018/entries/autonomy-moral/.
- Cudd, Ann E. 2006. *Analyzing Oppression*. Oxford: Oxford University Press.
- Dworkin, Gerald. 1988. *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press.
- Elster, Jon. 1983. "Sour Grapes." In *Sour Grapes: Studies in the Subversion of Rationality*, 110–41. Cambridge: Cambridge University Press.
- Enoch, David. 2006. "Agency, Shmagency: Why Normativity Won't Come from What Is Constitutive of Agency." *Philosophical Review* 115, no. 2: 169–98.
- Enoch, David. 2009. "Can There Be a Global, Interesting, Coherent Constructivism about Practical Reason?" *Philosophical Explorations* 12, no. 3: 319–39.
- Enoch, David. 2011a. "Giving Practical Reasons." *Philosophers' Imprint* 11, no. 4: 1–22.
- Enoch, David. 2011b. *Taking Morality Seriously: A Defense of Robust Realism*. Oxford: Oxford University Press.
- Enoch, David. 2013. "The Disorder of Public Reason." *Ethics* 124, no. 1: 141–76.
- Enoch, David. 2015. "Against Public Reason." In *Oxford Studies in Political Philosophy*, vol. 1, edited by David Sobel, Peter Vallentyne, and Steven Wall, 112–42. Oxford: Oxford University Press.
- Enoch, David. 2016. "What's Wrong with Paternalism: Autonomy, Belief, and Action." *Proceedings of the Aristotelian Society* 116, no. 1: 21–48.
- Enoch, David. 2017a. "Hypothetical Consent and the Value(s) of Autonomy." *Ethics* 128, no. 1: 6–36.

- Enoch, David. 2017b. "Political Philosophy and Epistemology: The Case of Public Reason." In *Oxford Studies in Political Philosophy*, vol. 3, edited by David Sobel, Peter Vallentyne, and Steven Wall, 132–65. Oxford: Oxford University Press.
- Frankfurt, Harry. 1971. "Freedom of the Will and the Concept of a Person." *Journal of Philosophy* 68, no. 1: 5–20.
- Hampton, Jean. 1993. "Selflessness and the Loss of Self." *Social Philosophy and Policy* 10, no. 1: 135–65.
- Healey, Richard. n.d. "Consent, Directed Duties, and Mutual Recognition." Unpublished manuscript.
- Hurd, Heidi. 1996. "The Moral Magic of Consent." *Legal Theory* 2, no. 2: 121–46.
- Kane, Robert. 2005. *A Contemporary Introduction to Free Will*. New York: Oxford University Press.
- Khader, Serene J. 2011. *Adaptive Preferences and Women's Empowerment*. Oxford: Oxford University Press.
- Killmister, Suzy. 2018. *Taking the Measure of Autonomy: A Four-Dimensional Theory of Self-Governance*. New York: Routledge.
- Lukes, Steven. 2005. *Power: A Radical View*. 2nd ed. London: Palgrave Macmillan.
- Mackenzie, Catriona, and Natalie Stoljar, eds. 2000. *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. New York: Oxford University Press.
- Meyers, Diana T. 1987. "Personal Autonomy and the Paradox of Feminine Socialization." *Journal of Philosophy* 84, no. 11: 619–28.
- Mill, John Stuart. (1859) 2011. *On Liberty*. Project Gutenberg. [www.gutenberg.org/files/34901-hhtml](http://www.gutenberg.org/files/34901/hhtml).
- Millgram, Elijah. 2000. "What's the Use of Utility?" *Philosophy and Public Affairs* 29, no. 2: 113–36.
- Mitchell, Polly. 2018. "Adaptive Preferences, Adapted Preferences." *Mind* 127, no. 508: 1003–25.
- Nozick, Robert. 1969. "Coercion." In *Philosophy, Science, and Method: Essays in Honor of Ernest Nagel*, edited by Sydney Morgenbesser, Patrick Suppes, and Morton G. White, 440–72. New York: Martin's Press.
- Nussbaum, Martha. 2001. "Adaptive Preferences and Women's Options." *Economics and Philosophy* 17, no. 1: 67–88.
- Oshana, Marina A. L. 1998. "Personal Autonomy and Society." *Journal of Social Philosophy* 29, no. 1: 81–102.
- Oshana, Marina A. L. 2006. *Personal Autonomy in Society*. New York: Routledge.
- Owens, David. 2012. *Shaping the Normative Landscape*. Oxford: Oxford University Press.
- Parfit, Derek. 2001. "Rationality and Reasons." In *Exploring Practical Philosophy: Essays in Honour of Ingmar Persson*, edited by Dan Egonsson, Jonas Josefsson, Björn Petersson, and Tony Rønnow-Rasmussen, 17–39. Farnham, UK: Ashgate.

- Railton, Peter. n.d. "Climate Change and Quality of Life." Unpublished manuscript.
- Railton, Peter. 1984. "Alienation, Consequentialism, and the Demands of Morality." *Philosophy and Public Affairs* 13, no. 2: 134–71.
- Raz, Joseph. 1986. *The Morality of Freedom*. Oxford: Oxford University Press.
- Richardson, Henry. 2001. "Autonomy's Many Normative Presuppositions." *American Philosophical Quarterly* 38, no. 3: 287–303.
- Schroeder, Mark. 2012. "The Ubiquity of State-Given Reasons." *Ethics* 122, no. 3: 457–88.
- Stoljar, Natalie. 2013. "Feminist Perspectives on Autonomy." Stanford Encyclopedia of Philosophy Archive, May 2. plato.stanford.edu/archives/win2015/entries/feminism-autonomy/.
- Stoljar, Natalie. 2014. "Autonomy and Adaptive Preference Formation." In *Autonomy, Oppression, and Gender*, edited by Andrea Veltman and Mark Piper, 227–52. Oxford: Oxford University Press.
- Stoljar, Natalie. 2018. "Answerability: A Condition of Autonomy or Moral Responsibility (or Both)?" In *Social Dimensions of Moral Responsibility*, edited by Katrina Hutchinson, Catriona Mackenzie, and Marina Oshana, 231–52. Oxford: Oxford University Press.
- Superson, Anita M. 2005. "Deformed Desires and Informed Desire Tests." *Hypatia* 20, no. 4: 109–26.
- Terlazzo, Rosa. 2016. "Conceptualizing Adaptive Preferences Respectfully: An Indirectly Substantive Account." *Journal of Political Philosophy* 24, no. 2: 206–26.
- Walker, John D. 1995. "Liberalism, Consent, and the Problem of Adaptive Preferences." *Social Theory and Practice* 21, no. 3: 457–71.
- Watson, Gary. 2009. "Promises, Reasons, and Normative Powers." In *Reasons for Action*, edited by David Sobel and Steven Wall, 155–78. Cambridge: Cambridge University Press.
- Weberman, David. 1997. "Liberal Democracy, Autonomy, and Ideology Critique." *Social Theory and Practice* 23, no. 2: 205–33.
- Westlund, Andrea. 2003. "Selflessness and Responsibility for Self: Is Deference Compatible with Autonomy?" *Philosophical Review* 112, no. 4: 483–523.
- Williams, Bernard. 2009. *In the Beginning Was the Deed: Realism and Moralism in Political Argument*. Princeton, NJ: Princeton University Press.