9 Autonomy Under Oppression
Tensions, Trade-Offs, and Resistance

Suzy Killmister

1. INTRODUCTION

It is a commonplace in philosophical discussions of oppression to assume that it compromises personal autonomy. Indeed, this is often taken to be one of oppression's most distinctive and insidious harms. My aim in this chapter is to explore precisely how oppression compromises personal autonomy and what individuals might be able to do to insulate themselves from these effects. The focus of my discussion will be on the distinctive harms of double binds. Double binds occur when an agent is faced with a decision, either side of which subjects her to some form of sanction or frustration. In this respect, they are analogous to moral dilemmas. A classic example of a double bind is the situation faced by many women in contemporary Western societies with respect to career and family: given the absence of adequate child-care options, many women must still choose between fulfilling their career goals and raising a family. I argue that double binds such as these compromise personal autonomy: by their very nature, such double binds prevent individuals from realizing some or other goal or value. Moreover, given the inescapability of choice, these double binds force individuals to actively frustrate one of their goals or values, which, as I will go on to show, necessarily constitutes a reduction in the agent's autonomy.

At this point a curious problem emerges: if, as I will go on to argue, autonomy is compromised because of mutually unrealizable goals and values, then it looks like personal autonomy could flourish under oppression if the agent successfully undertook a project of self-modification. In other words, if the agent molded her values and goals to fit the demands of the oppressive society, then the conflict that threatened her autonomy would be avoided. I will concede that insofar as this process is viable, some individuals may be better able to protect their autonomy by capitulating to oppression than by resisting it. However, I suggest that the extent of modification required to both avoid double binds and preserve autonomy is psychologically implausible. For the most part, such modifications will fall short, and when they do, they merely relocate the reduction in autonomy from one dimension of autonomy to another (what these dimensions are will be outlined shortly).
using “unreasonableness” as an umbrella term to cover any standard the agent might appeal to in explaining why she finds the experience of a particular attitude to be problematic. For instance, an agent may find herself frightened of a puppy and be baffled by her own response. In explaining her bafflement, she might point to such things as the puppy being obviously unthreatening. Such a response would suggest that this agent takes something like fitness to be a relevant standard for assessing the reasonableness of fears (at least in some contexts). By contrast, another agent may find herself sexually attracted to a particular individual but deem that attraction to be morally inappropriate (perhaps because that individual is her best friend’s partner). What is important to stress here is that the idea of reasonableness is subjective in two distinct senses. First, the standard by which an attitude is judged (i.e., fitness or moral appropriateness) is up to the agent herself. If an agent doesn’t take moral inappropriateness to compromise motivational attitudes, then it doesn’t affect her Self-Definition if she experiences a motivational attitude that she takes to be morally inappropriate. Second, whether or not the motivational attitude meets whatever standard the agent takes to be relevant is also up to the agent. So even if, objectively, fear of a puppy should be considered unfruitful, if the agent herself considers the fear fitting (and hence reasonable), then her fear of the puppy makes her no less Self-Defined.

To return to the example of sexual oppression, if an agent in a deeply homophobic society comes to believe that her homosexual desires are deviant, as she is relentlessly socialized to do, then she will very likely be led to reject them as unreasonable. Her society offers her a moral standard by which to judge her desires, and if she internalizes that standard as appropriate, then the judgment of unreasonableness follows as a matter of course. For such an agent, to experience homosexual desires will conflict with how she takes herself to have reason to be and will thus constitute a reduction in her autonomy.

2.2. External Self-Realization

A second way in which oppression typically compromises autonomy is through frustration of External Self-Realization. According to External Self-Realization, if an agent intends to A but in fact B, she is failing to fully govern herself. One of the paradigmatic ways in which External Self-Realization can be frustrated is through an external agent’s intervention in the realization of an intention. For instance, if you have formed an intention to eat a cookie, but I physically prevent you from eating it (perhaps reasoning that you can’t really want it), then I have frustrated your External Self-Realization. By substituting my will for yours, I have constrained your relation to yourself in the same way as the captain moored to the dock has a constrained relation to her ship: your ability to issue directives to yourself may remain untouched, but insofar as those directives are blocked from receiving uptake in action, they cannot fully secure your self-governance.

Under conditions of severe oppression, an agent simply may not have the option to convert her intention into the corresponding action. For instance, take a Palestinian youth in the Occupied Territories. Her intention to visit a friend in Israel is subject to the approval of the checkpoint guard; if he refuses her access, her intention cannot be realized and her External Self-Realization is thus frustrated. Or consider a woman in an abusive relationship. Though she may repeatedly form an intention to leave, her fear of the consequences of following through on that intention, for both herself and her children, may prevent her from translating that intention into action.

2.3. Internal Self-Realization

A third way in which oppression commonly compromises autonomy, and the primary focus of this chapter, is through the creation of double binds. Double binds, I suggest, inevitably reduce an agent’s Internal Self-Realization through the presence of reasons that cannot be fulfilled. The greater the extent to which those unrealizable reasons reflect core values or commitments of the agent, the greater the extent to which the agent’s autonomy is reduced. As such, the greater the extent to which a double bind creates a dilemma between two or more of the agent’s core values or commitments, the greater the extent to which that double bind reduces the agent’s autonomy.

As noted, Internal Self-Realization measures the extent to which the agent’s intentions map onto the reasons she takes herself to have. Akrasia would be a paradigmatic case of reduced Internal Self-Realization: the agent judges that she has most reason to A, but she nonetheless forms the intention to ¬A. Insofar as the agent’s intention comes apart from her practical judgment, she is failing to realize her commitments—she is failing to fully govern herself.

The crucial question for Internal Self-Realization is what it means for an agent to take something to be a reason for action. While I do not think that reasons for action necessarily reduce to motivational attitudes, for simplicity’s sake I will consider desires as paradigmatic candidates for reasons for action in what follows. On a simple model, weighing reasons would simply amount to weighing up one’s desires; the stronger my desire for x, the more reason I have to pursue x. Building on Scanlon, however, I suggest that not all desires are necessarily taken up as reasons for action in an agent’s deliberation; they may instead be “bracketed” as reasons. Even when they are taken up as reasons for action, moreover, their strength as a reason may not necessarily correspond to their motivational strength. For an example of the latter case, we can consider an agent deliberating about whether to continue watching TV or instead to go to the gym. Whether her ensuing intention to
continue watching TV exhibits high or low Internal Self-Realization depends not on whether her desire to watch TV is stronger than her desire to go to the gym but instead whether the agent takes her desire to watch TV to provide her with greater reason than her desire to go to the gym.

Bracketing is a more complex phenomenon, but the following example illustrates the point. Imagine experiencing a sudden desire to eat a Big Mac despite not being remotely hungry or even particularly enjoying the taste of Big Macs. You know, moreover, that the only explanation for your sudden desire is that you have just seen a billboard advertising a Big Mac. In such a case, you may take your desire for a Big Mac to provide you with no reason whatsoever to stop at MacDonald’s. It is not that this desire is outweighed—perhaps by concerns for your health—but rather that the desire is denied any weight in your consideration of what to do.

If an agent acts solely on the basis of a desire that she has bracketed, then by her own lights she is acting without reason. This would involve a serious reduction of Internal Self-Realization. Less damaging but still problematic are cases in which an agent forms an intention to fulfill an unbracketed desire that she takes to be outweighed by other reasons. While the intention is suboptimal as an expression of the agent’s will, it at least finds some traction in her calculus of reasons. Finally, and perhaps less intuitively, if an agent forms an intention that corresponds with what she takes herself to have most reason to do, she will still experience a reduction in Internal Self-Realization insofar as that intention fails to realize outweighed but unbracketed desires. For example, if on the earlier example I take myself to have more reason to watch TV and thus form that intention, the fact that my reasons for going to the gym remain unrealized reduces my Internal Self-Realization. As noted, typically an agent’s core values and commitments will constitute stronger reasons than mere inclinations, meaning that to assess the degree of harm to Internal Self-Realization of an unfulfilled reason, it will be helpful to know whether that reason was tracking a value or commitment, on the one hand, or simply an endorsed whim, on the other.11

Double binds, by their very nature, limit the possibility for an agent to realize all the reasons that she takes herself to have. To consider a mundane example, take a woman in contemporary U.S. society deciding whether to shave her legs. On the one hand, she takes herself to have strong reasons not to: she considers it an oppressive patriarchal norm; not to do so would set a good example to younger generations; shaving one’s legs takes time and money better spent on other activities. On the other hand, she recognizes strong reasons to capitulate: precisely because it’s a patriarchal norm, she will be subject to varying degrees of social discipline if she presents herself au naturel; moreover, having been socialized into Western beauty norms, she knows she will struggle with a visceral aversion to the way she looks unshaven. The structure of this dilemma is common enough. One of the key characteristics of oppression is that it creates disincentives for members of the oppressed group, which give them reasons not to do that which the societal norms deem inappropriate for agents like them.12 Examples abound: women trying to balance a career with the desire to have a family; gay people deciding whether to come out to friends and family in the face of homophobia; people of color hoping to pursue a career in academia despite institutional racism. Whatever path the agent takes under such double binds, she will be falling short of doing all that she takes herself to have reason to do; some of her reasons will go unrealized in her action, and some of those reasons may well reflect core values or commitments. So if our original agent decides to shave her legs, she will be thwarting her commitment to making a stand against patriarchy; if she decides not to, she will be thwarting her value of experiencing a positive self-image. While in neither of these scenarios is she thereby non-autonomous, the requirement to “bargain with patriarchy”13 reduces her autonomy by restricting the extent to which her intentions can realize all of her reasons for action.

In situations of double binds, I am suggesting, the agent is less autonomous than she would be if she experienced no countervailing reasons, because even outweighed reasons remain reasons for action; insofar as her action frustrates any of her reasons for action, she is less autonomous than she might otherwise be.14 This claim may strike some as counterintuitive and so requires further comment. When our reasons for action conflict, I believe that it is a mistake to too strongly identify the self with the “winning” side. If I am torn between protecting my career and caring for my children, both of these motivations speak for me because both reflect a core value. Whichever decision I make, some important part of me is frustrated, and so that choice does less to realize my self than it might otherwise, such as if there were a path available that allowed both values to be fulfilled simultaneously.15 If there were a way for these values to both be fulfilled, my action could more fully capture what I took myself to have reason to do and thus could more closely express my will. An action that fulfilled everything I took myself to have reason to do would be closer to the ideal of self-governance, whereby what I do and who I am fully reflects my reasons and commitments.

It may be objected that there are obvious counterexamples to this claim. For instance, take two agents each deciding whether to order tea or coffee at a restaurant. Both prefer coffee, but one dislikes tea while the other likes it. Surely we do not want to have to say that the agent who likes tea but prefers coffee has her autonomy reduced by her conflicting desires. However, this kind of conflict is importantly disanalogous to double binds. In these standard choices, the reason we have for either option is in fact the same reason: in this case, to drink a tasty beverage. Once I have settled on which tasty beverage I prefer, that reason is fulfilled, and I no longer have a reason to drink tea.16 This is manifestly not the case in all exclusive choices, particularly when the agent’s core motivational attitudes are at stake. To make this especially vivid, consider an agent confronting a particularly tragic dilemma such as Sophie in the eponymous book Sophie’s Choice.17 When the concentration camp guard confronts Sophie with the tragic decision of which
child to sacrifice, he thereby reduces her autonomy, because he makes it impossible for her to act in accordance with what she takes herself to have profound reason to do. Her desire to protect her daughter does not cease to be a reason for her just because, in that instant, she has a stronger desire to protect her son. Or, to put it slightly differently, her reason to protect her daughter is not the same reason as her reason to protect her son; doing the latter does not fulfill the former. The same holds for the vast proportion of coercive threats: even though the agent does what she most wants to do, given the threat situation, the defeated desires continue to function as reasons, and their frustration thus reduces the agent's autonomy. Precisely what is wrong with coercive threats is that they use some of the agent's motivational attitudes against herself—they rely on the strength of the desire to survive, avoid pain, and so on, to frustrate other of the agent's goals. If we suppose that the failure to fulfill conflicting reasons does not reduce autonomy, we lose the ability to explain why coercive threats reduce autonomy. Once the autonomy-reducing nature of coercive threats is understood, however, we can see that the very same dynamic is in play in double binds.  

3. SELF-MODIFICATION TO PRESERVE AUTONOMY

So much for the ways in which autonomy is commonly reduced under conditions of oppression. I want to turn now to consider the possibility that—for at least for some agents, some of the time—capitulating to oppression can augment autonomy. The motivation for this suggestion is simple: if the problem with double binds is that they leave some of the agent’s values or goals unfulfilled, then the problem would dissolve were the agent to bring her values and goals into alignment with the oppressive norms. To put it slightly differently, an agent whose motivational attitudes, values, and goals wholeheartedly reflected the norms and expectations of society would not face double binds. What we see here is essentially a Stoicist move: it doesn’t matter if society only gives me limited space to move, provided I can shrink myself to fit the space that I am given. This is what Isaiah Berlin has famously characterized as a “retreat to the inner citadel.” Since on the theory that I have put forward autonomy is constituted by satisfaction with, and realization of, our motivational attitudes and commitments, it is at least conceptually possible that autonomy could be augmented by the kind of “retreat to the inner citadel” that so concerns Berlin. I might become more autonomous by capitulating to oppression.

While conceptually possible, I will argue that such a move is nonetheless implausible. There are two reasons for such implausibility. First, only some kinds of double binds can be resolved through aligning one’s attitudes with oppressive norms. Second, even when aligning one’s motivational attitudes with oppressive norms does dissolve the conflict inherent in double binds, it commonly does so merely by pushing the problem onto another dimension of autonomy. An agent making this move is thus making a trade-off: she is augmenting one aspect of her autonomy at the price of another.

The first problem with the claim that autonomy could be protected through the cultivation of different motivational attitudes is that not all double binds are of the form presented in the leg-shaving example. In that case, the bind was primarily caused by a clash between the agent’s values and the oppressive norms. If the agent aligned her values with the oppressive norms, her behavior would presumably receive no sanction, and thus she would not face any impediment to the realization of those values. Compare this to the kind of double bind that interests Marilyn Frye, whereby whatever the agent does, she is subject to sanction and censure; double binds of this type “expose one to penalty, loss or contempt whether one works outside the home or not, is on welfare or not, bears children or not.” For a woman in a deeply patriarchal society, cultivating motivational attitudes that favor working in the home or bearing children won’t necessarily help, because even if the agent succeeds in realizing those now internalized values, she will nonetheless be confronted with social contempt in view of the low esteem placed on what are seen as traditional women’s roles. Nor could she escape censure by cultivating motivational attitudes for working outside the home or remaining childless: such roles would also induce social contempt in virtue of the violation of proscribed gender roles. Assuming she places some value on self-respect, any way she acts will frustrate that value.

Let’s return now to the kind of double bind that may initially look like it could be resolved through the cultivation of different motivational attitudes, such as the woman deciding whether to comport with oppressive standards of beauty. In contrast to Frye’s examples, we may suppose that in such cases, the woman would face no censure or sanction for comporting with those standards; it is not the case that, whatever the agent does, she is subject to sanction and censure. Instead, all that stands in her way are her own desires and values, which provide her reasons for not comporting with the oppressive standards. There are at least two avenues such an agent might pursue, were her only concern to augment her autonomy. On the one hand, she could try to erase the desires and values that provided the reasons not to comport with the oppressive standards. Conversely, she could tolerate the desires and values while rejecting them as reason giving. To explain this latter move: since reductions of Internal Self-Realization only occur when the agent fails to realize unbracketed desires, the agent’s Internal Self-Realization would be unaffected by failure to realize desires that she did not take to be reason giving.

While the latter strategy may appear more psychologically plausible, I want to suggest that it is nonetheless deeply problematic. The problem is that this strategy merely displaces the problem from Internal Self-Realization to Self-Definition, and this is because to reject the relevant desires in this case as reason giving, as opposed to simply taking them to be outweighed, involves taking them to be false. It is important to be clear that this is not
necessarily the case for all bracketed desires. As Scanlon explains, desires are commonly bracketed when we engage in moral reasoning; an agent may well bracket a self-interested desire when it conflicts with moral requirements, without rejecting that desire in other contexts. However, in those kinds of cases, we have a ready story for why the desire in question is only conditionally reason giving, and this story necessarily invokes the role of morality in deliberation. By contrast, what story could we possibly tell ourselves about why our desire to resist oppression should be bracketed, as opposed to merely being outweighed? Because such a desire is already in the moral domain, I suggest, it can only plausibly be bracketed by being deemed irrelevant to the situation. We would have to assume that the relevant beauty norms were not oppressive, or that such oppression does not call for personal resistance. Such an assumption, however, would be exceptionally difficult to reconcile with the judgment that the desires to resist those norms were reasonable. In other words, assuming that a norm is not oppressive would typically be accompanied by the judgment that a desire to resist that norm is, ceteris paribus, unreasonable. Any emotional resistance to oppressive beauty norms would thus come to be seen as misguided; the agent would be subject to motivational attitudes that she deemed unreasonable. What this means, then, is that typically an agent can only manipulate her motivational attitudes to protect her Internal Self-Realization at the cost of reducing her Self-Definition.

This leaves the possibility that the agent could simply rid herself of her resisting desires. If this were possible, there would be no motivational attitudes that the agent deemed unreasonable, and hence no reductions in Self-Definition. Likewise, there would be no reason-giving desires that failed to be realized in her intentions and actions, and hence no reduction in either Internal or External Self-Realization. Such an agent would augment her autonomy through careful manipulation of her motivational attitudes, weeding out those in conflict with either social norms or other of her attitudes. She would augment her autonomy by, in Berlin’s phrase, retreating to the inner citadel.

I am accepting for the purposes of this chapter that such projects of self-modification are possible. If adaptive preferences are a real phenomenon, as I am assuming they are, then desires are malleable, particularly in the face of social pressure. Since the self-modification we are imagining is toward a self that coheres with social demands, it can piggyback on the multiplicity of socialization practices that are already urging her in that direction.

At this point, a simple objection may suggest itself, namely that motivational attitudes induced in such a way are by definition non-autonomous. Surely, we might think, there’s a difference between an agent who has cultivated motivational attitudes under conditions of freedom and one who has cultivated such motivational attitudes in response to oppressive conditions. Unfortunately, this simple objection does not go through. To appeal to the fact that certain kinds of socialization are qualitatively different in that they are oppressive is simply to push the problem back a step. What is it about oppressive socialization that is supposed to compromise autonomy? If the answer is simply “that it is oppressive,” it looks worryingly ad hoc. What’s needed is an explanation as to how, by virtue of their origins, oppressive norms are necessarily heteronomous in a way that other norms into which we’re socialized are not. Yet if there is nothing within the agent’s psychology that resists these norms, then it is hard to see where the accusation of heteronomy is supposed to gain traction.

It may be objected that lack of resistance is insufficient, since that lack is due to precisely the forces that we are worried about as autonomy undermining. Yet attempting to accommodate this worry seems to inevitably cast the net too wide. For instance, if we try to explain what’s wrong with oppressive norms in terms of whether the agent would approve of the new norms in advance of the modification, we seem to rule out cases of radical transformations. When an agent undergoes a religious conversion or some other kind of transformative epiphany, it is commonly true both that the agent’s earlier self would not approve of the result of the transformation and that she would not approve of its means. Just as for the oppressed agent, the religious convert’s lack of resistance to her new motivational attitudes is explicable entirely in terms of a transformative process that she neither chose nor preemptively endorsed. Unless we want to say that in all such cases the posttransformation agent is heteronomous, we need to concede that antecedently disapproving of the later motivational attitudes is insufficient to label those attitudes non-autonomous.

There is a grain of truth, though, in the worry that a wholesale retreat to the inner citadel can undermine autonomy. To see why this is so, it will be useful to revisit and extend Berlin’s political analogy. If a government under siege were to cede all of its territory, retreating to an underground bunker but continuing to send unheeded orders into the outside world, we would rightly mock any claims of self-governance. There is simply no realm left to govern. The same phenomenon can occur at the individual level. My resistance to following the standard line on adaptive preferences lies in a concern to index autonomy to the agent’s self, and in particular to her reflectively endorsed motivational attitudes and commitments. Even if that self is impoverished, in that it is composed of a limited, socially sanctioned set of values, the agent can still be self-governing, since she has the self she wants to have, and that self is effectively governing her actions. She is acting in accordance with her commitments. However, there is a point at which impoverishment of the self turns into self-effacement; and without a minimal self, there can be no self-governance.

To be self-governing, the agent must be able to acknowledge at least some of the motivational attitudes she experiences as her own and, more importantly, she must be able to accept at least some of them as reason giving. Otherwise, the self loses all capacity to govern—there is nothing there from which a direction might issue. Consider Paul Benson’s example of the
gaslighted woman: an agent becomes convinced of her own insanity, such that she no longer trusts herself to competently self-govern. In such cases, arguably, the agent becomes fully estranged from her motivational attitudes, such that they are no longer even candidates as reasons for action. Such extreme cases amount to a kind of erasure of the self. It is like a government trying to rule in the absence of a realm. While such an outcome remains a possibility under conditions of severe oppression, it is important not to overstate the case: in most cases, oppression is insufficient to bring about such complete erasure of the self, and insofar as there remains a self, we cannot assume that the agent is incapable of self-governance. There is, however, an additional reason to suspect that the project of modifying one's motivational attitudes might not be as conducive to autonomy as this response might suggest. To see why, though, requires a clarification of the conditions of autonomy.

As I have sketched Self-Definition and Internal Self-Realization, I have been concerned only with the actual judgment of the agent concerning the reasonableness and reason-giving force of her motivational attitudes (which we might think of as two varieties of endorsement). However, as the extensive literature on this issue attests, actual endorsement—however it is construed—leaves the door open to cases in which that endorsement is secured via various modes of manipulation. If the only reason I take my fear of puppies to be reasonable is because I have been interfered with by a nefarious brain surgeon or brainwashed by a cult leader, it would seem strange to say that I am nonetheless autonomous. Similarly, if I take myself to have most reason to drink the contents of the glass because, and only because, I have been hypnotized to think the glass contains water, when in fact it contains petrol, it would seem strange to say that in drinking the contents of the glass, I am acting autonomously. Clearly, a modification is called for.

The required modification takes the agent's endorsement to be necessary but not sufficient to secure autonomy; full autonomy requires both endorsement—in both the "reasonableness" and "reasons for action" senses outlined—and an additional condition, which I will call the "consistency condition." This consistency condition builds on the work of John Christman, though in terms that stray quite far from his theory. For Christman, what matters is whether the agent would revoke her endorsement for (or, in his terms, feel alienated from) her desire were she to know of its history. I see no reason to stop with the introduction of this particular realm of facts, however. We could similarly ask, and for the same reasons, whether the agent would revoke her endorsement of a desire were she to know pertinent facts about its object—for instance, that the person she is attracted to is in fact her best friend's partner.

This raises the difficult question of which facts to introduce. The danger is that, in requiring endorsement to survive the introduction of relevant facts, autonomy will collapse into orthodoxy. Why should it matter that an agent acts on the basis of a misguided norm if that norm is what she is truly committed to? What is needed is a way to ensure that the facts that are introduced are relevant to the agent's autonomy and not merely to the objective correctness of her action or attitude. Moreover, putting the modification in terms of countermotional endorsement introduces all the problems that typically attend counterfactuals and is thus best avoided.

To gain around these problems, I propose the consistency condition. It matters for an agent's autonomy whether her endorsement is consistent with the vast web of epistemic and practical commitments that she otherwise holds, because failure to uphold any of these epistemic or practical commitments is a failure to act in accordance with one's own rules—it is a failure of self-governance. So in the hypnotism case, I take myself to have a reason to drink the contents of the glass; but I also have a practical commitment to acting in light of how the world actually is. This makes the content of the glass relevant to my determination of my reasons for action and hence a fact that needs to be taken into consideration. As such, we need to ask whether I am currently disposed to take myself to have a reason to drink the contents of the glass, given that it is petrol. If the answer is no, then my intention to drink the contents of the glass does not change what I am committed to taking myself to have reason to do and so that fulfilling that intention does not raise my Internal Self-Realization. A similar story can be told about the puppy case. I take my fear to be reasonable; but I also have a background commitment to fear being sensitive to fitness (at least in these kinds of cases) and to nondangerous things being unfitting (again, in cases relevantly similar to the one at hand). Since there is an inconsistency between my judgment that my fear is reasonable and what I am committed to accepting as reasonable, that fear reduces my level of Self-Definition.

What this means in practice is that it would be insufficient for an agent to strip away the motivational attitudes that directly conflict with oppressive norms; to augment her autonomy, she would also have to strip away all of the epistemic and practical commitments that were inconsistent with her endorsement of her new motivational set. Our agent confronted with patriarchal beauty norms would thus have to come to desire to uphold those norms and rid herself of any commitment that conflicted with endorsing those desires. This would require a wholesale reconfiguration of her orientation to the world, potentially involving rooting out any commitments she might have to justice, equality, and truth.

Just as attempts to rescue Internal Self-Realization by changing what we take ourselves to have reason to do tend to simply relocate the problem to Self-Definition, so too attempts to rescue Self-Definition by changing our motivational attitudes tend to simply relocate the problem to the consistency condition. The agent concerned for her autonomy will thus be reduced to simply making trade-offs within the different dimensions of autonomy. Nonetheless, it would be too strong to say that wholesale modifications are actually impossible. As I will argue in the final section, however, while such
If one's motivation to abandon a commitment is that doing so will make it easier to live in an unjust society, the modification may be understandable, but it involves a loss of integrity.

To be clear, the kind of integrity-damaging modification I am imagining is compatible, at least conceptually, with fully maintaining one's autonomy. On the account sketched earlier, autonomy requires that one judges one's motivational attitudes to be reasonable, and that one forms and enacts intentions that reflect what one takes oneself to have reason to do. If one were to successfully modify one's motivational attitudes in order to make it easier to live in an unjust society, along with the web of epistemic and practical commitments with which these attitudes and their cultivation must cohere, then one's autonomy would remain intact.33

The upshot of this is that the kinds of modifications that seek to preserve autonomy under oppression necessarily undermine integrity, and this is so whether or not those modifications are successful. The modifications are undertaken as a response to the pressure created by oppressive circumstances, and as such, they are incompatible with maintaining one's integrity. Agents under oppression thus face yet another tragic choice: if they aim at preserving their autonomy, they must accept the loss of their integrity.

5. CONCLUSION

I have argued that oppression compromises autonomy through the generation of double binds. Under double binds, agents are unable to fully realize their goals, values, and commitments and are thus less than fully autonomous. While such agents retain some control over their own autonomy, this control largely amounts to the power to undergo trade-offs. Agents can choose to protect their External Self-Realization by only developing intentions that are possible to realize, but this reduces their Internal Self-Realization; they can undergo a process of modifying what they take themselves to have reason to do in order to protect their Internal Self-Realization, but then problems emerge for their Self-Definition; they can try to strip away all of the motivational attitudes that conflict with the oppressive norms, but this merely pushes the problem onto the consistency requirement. Finally, they could attempt to undergo a wholesale reorientation of their epistemic and practical commitment, but to do so would be to abandon their integrity. Oppression thus presents the oppressed with a particularly egregious type of tragic dilemma: whatever they do, something of value must be sacrificed.

Marilyn Frye has described oppression as like a cage. Focusing on any single bar makes it seem like it should be possible to escape, but this promise disintegrates once the cage is viewed in its entirety. For an agent concerned with protecting her autonomy under oppression, there are many bars that appear circumventable. Circumventing any one, however, merely runs her headlong into a neighboring bar. Agents under oppression may well have the
freedom to choose which aspect of their autonomy to sacrifice or whether to protect their autonomy at the price of their integrity, and this is better than no freedom at all. But it is still just the freedom of the caged bird to choose which bar impedes her flight.

ACKNOWLEDGMENTS

Many thanks to Marina Oshana and Natalie Stoljar for helpful comments on earlier drafts and to the audience at the Workshop on the Duty to Resist One’s Own Oppression, University of Connecticut, 2014, where this chapter was first presented.

NOTES


2. See Cudd, Analyzing Oppression.

3. This claim goes against the grain of much of the literature on oppression, whereby adapting one’s desires to align with oppressive norms is considered necessarily autonomy undermining. For a variety of arguments supporting the standard view, see, e.g., Cudd, Analyzing Oppression; Jon Elster, Sour Grapes: Studies in the Subversion of Rationality (Cambridge: Cambridge University Press, 1985); Martha Nussbaum, Women and Human Development: The Capabilities Approach (Cambridge: Cambridge, 2001); Superson, “Deformed Desires and Informed Desire Tests”; Natalie Stoljar, “Autonomy and Adaptive Preference Formation,” in Autonomy, Oppression and Gender, ed. Andrea Velten and Mark Piper (New York: Oxford University Press, 2014), 227–252.

4. I take these reasons and commitments to constitute the relevant notion of “self” being appealed to in the idea of “self-governance” in that they form the standpoint with the authority to speak for the agent (though I take no stance here on broader metaphysical questions of the self, such as identity over time). This theory thus follows in the tradition of the second-order dualistic theories of autonomy first developed by Harry Frankfurt and Gerald Dworkin and has much in common with Michael Bratman’s theory of self-governance. See Harry Frankfurt, The Importance of What We Care About (Cambridge: Cambridge University Press, 1988); Gerald Dworkin, The Theory and Practice of Autonomy (Cambridge: Cambridge University Press, 1988); Michael Bratman, Structures of Agency: Essays (Oxford: Oxford University Press, 2007). For a discussion of how and why my theory diverges from these, see Suzy Killmister, “The Woody Allen Puzzle: How ‘Authentic Alienation’ Complicates Autonomy,” Noûs 10 (1111/ nous.12069); The theory is developed in more detail in Suzy Killmister, Taking

Autonomy Under Oppression

5. As will become clear, my condition of Self-Definition differs significantly from Diana Meyers’s condition of the same name. For Meyers, self-definition involves bringing a set of agency skills to bear on questions such as “What sort of person am I?” and “What really matters to me?” See Diana Meyers, Self, Society, and Personal Choice (New York: Columbia University Press, 1989); Gender in the Mirror: Cultural Imagery and Women’s Agency (New York: Oxford University Press, 2002).

6. To be a maximally autonomous agent, on this theory, each dimension would need to be fully realized. I am assuming that such a feat is beyond the capabilities of all human persons. As such, this theory is not intended to directly set a threshold for what counts as “autonomous” but instead to provide a schema for analyzing the various ways in which autonomy can be compromised.


8. I should stress that frustration of External Self-Realization may not be sufficient for an all-things-considered judgment of non-autonomy: an agent’s level of autonomy is a factor of all three dimensions. It is part of my broader approach that different applications of autonomy (i.e., as part of moral responsibility or as a basis for valid consent) will privilege different dimensions, so that the same agent may be judged autonomous in one respect but not in another.


11. The extent to which Internal Self-Realization is reduced is a product of the strength of the reason that is unrealized, where for most agents this will in turn be a product of that reason’s relationship to their core values and commitments, the more central a value or commitment is to the agent, the greater the extent to which its failure to be realized reduces the agent’s autonomy. On my account, then, even trivial frustrations will technically reduce an agent’s Internal Self-Realization, but for virtually any purpose for which we’re concerned with the level of an agent’s autonomy, these reductions will prove inconsequential.


15. As this example may suggest, oppositional practices are not the only way in which such tragic conflicts can arise. Sarre’s example of the young man choosing between fighting for his country and staying with his ailing mother has precisely the same shape and so, on my view, will be equally damaging to his autonomy as cases that arise through oppression. What’s distinctive about oppression is not that it affects autonomy but that it does so in ways that disproportionately target disadvantaged groups, and is politically imposed rather than inevitable.


18. This is not to suggest that all double binds reduce autonomy to the same extent. What matters for measuring Internal Self-Realization is the weight we give to the reasons that we fail to realize. On this measure, the extent to which Sophie's autonomy is reduced by the double bind she faces is exponentially greater than the extent to which the leg shaver's autonomy is reduced by her double bind.


21. This final assumption could be challenged: if such an agent were to cultivate a desire to be treated with contempt, then adopting traditional gender roles could bring about the full realization of her goals and values. This suggests that another trade-off may be on the horizon, this time between autonomy and self-respect. Space does not allow for exploration of this possibility here, but cf. John Rawls' claim that the social bases of self-respect are a primary good, *A Theory of Justice* (Oxford: Oxford University Press, 1999).

22. For a more detailed discussion of the relationship between treating an attitude as reasonable and treating it as a reason for action, see Killmister, "The Woody Allen Paradox.


26. This suggests that a minimal degree of self-trust is necessary for Internal Self-Realization. Arguably, it is also necessary for Self-Definition, insofar as the agent needs some degree of self-trust to generate judgments of reasonableness, and for External Self-Realization, insofar as the agent needs some degree of self-trust to generate the confidence to convert her intention into action. This requirement for self-trust should be distinguished from Andrea Westlund's requirement that the agent have a disposition to answer for herself. For a critique of Westlund's view, see Suzy Killmister, "Autonomy and the Problem of Socialization," *Social Theory and Practice* 39 (2013), 95–119.


31. I take it that commitments don't have to be consciously held, they can be revealed through our deliberative and practical histories.


33. More accurately, we might say that one would end up with a higher level of autonomy. Presumably one's autonomy would take a hit during the process of cultivation, before the conflicting commitments had been weeded out.

REFERENCES


